TOWARDS WIRELESS VIRTUALIZATION FOR 5G CELLULAR SYSTEMS

A Thesis
Presented to
The Academic Faculty

by

Albert Gran Alcoz

In Partial Fulfillment of the Requirements for the
Bachelor's Degree in Science and Telecommunication Technologies Engineering in the
Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona
Universitat Politècnica de Catalunya

Georgia Institute of Technology
July 2016

TOWARDS WIRELESS VIRTUALIZATION FOR 5G CELLULAR SYSTEMS

Approved by:

Professor Ian F. Akyildiz, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Ferran Casadevall, Advisor
School of Telecommunication Engineering
Universitat Politècnica de Catalunya

Date Approved: 14 July 2016

To my parents

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to Professor Akyildiz, for having given me the opportunity and the privilege of achieving one of my biggest dreams, for having appreciated all my efforts, for having trusted in my job and for all the support and inspiration during these six months.
I would also like to thank Eduard Alarcón, for having believed in me for this position and for having convinced me that with illusion and hard work everything that you desire can be achieved. To the best lab mates: Abdallah, Carlos, Bige, Ozan, Ayse, Manoj, Shuai and Chong, Can and Ibrahim and my mentor Shih-Chun, for having opened me the path to the research world and for all the wise words and advices towards my professional career. To Professor Casadevall for all the guidance and dedicated involvement in every step throughout the project. And to all my professors of Barcelona, for teaching me everything I know in the field, when I still did not know how helpful would it be.
Getting through a new stage of your life requires much more than academic support. I would like to thank all these people without whom this experience would not have been the same. Roser, Pol, Ricardo, Cusuh, Stefano, Ignacio, Megi, Parsa, you have been my family in Atlanta and you will always be. To all the amazing people from University House, iHouse and SQ5 and to Joe, José Luiz, Van and Zack, thank you for being the best roommates I could have ever asked for.
Finally to all my family and friends that have made me feel like at home even being 8000 km away. And to the two most important people in my life, my parents, no hi ha paraules que expressin tot el que us estimo. Aquesta tesis és per vosaltres.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Although it has been defined as one of the most promising key enabling technologies for the forthcoming fifth generation cellular networks, wireless virtualization still has several challenges remaining to be addressed. Amongst those, resource allocation, which decides how to embed the different wireless virtual networks on the physical relying infrastructure, is the one receiving maximum attention. This project aims at finding the optimal resource allocation for each virtual network, in terms of channel resources, power levels and radio access technologies so that the data rate requested by each virtual network can be guaranteed and the global throughput efficiency can be maximized.

CHAPTER I

INTRODUCTION

The next generation of cellular wireless networks, also known as 5G, is envisioned to provide higher data rate, lower end-to-end latency, improved spectrum and energy efficiency and reduced cost per bit [2]. With it, some concepts such as Infrastructure as a Service will become more and more familiar, with the idea of having different mobile virtual network operators providing distinguished services to their subscribed users and sharing a common infrastructure becoming a reality. This brings up the necessity of dividing the physical limited resources between the virtual networks to accommodate the dynamic demands of the continuously increasing and highly variable traffic patterns of their subscribed users, while satisfying the virtualization requirements of efficient resource allocation, intra-slice customization and inter-slice isolation.

The objective of this thesis is to move towards the design and implementation of optimal algorithms for wireless virtualization in a way that the QoS requirements of the different virtual networks can be satisfied and the global throughput efficiency can be maximized. In this first point of the thesis the main ideas that surround wireless virtualization will be discussed, as well as the key design issues that will need to be taken into account throughout all the project. The state of the art will be reviewed and the working environment scenario will be exposed to start moving towards the implementation of our proposed solutions and their respect algorithms.

1.1 Virtualization in the Wireless Domain

Virtualization has been seen as one of the main evolution trends in the forthcoming fifth generation (5G) cellular networks [15]. It involves abstraction and sharing of resources among different parties. With virtualization, the overall cost of equipment and management can be significantly reduced due to the increased hardware utilization, decoupled functionalities from infrastructure, easier migration to newer services and products and flexible management [8].

In wired networks, virtualization has been exploited during the last years, being considered one of the most promising technologies for the future internet. With the tremendous growth in wireless traffic and services, it is natural to extend virtualization to wireless networks.

With wireless network virtualization, the network infrastructure is decoupled from the services it provides. Multiple wireless virtual networks, operated by different service providers (SPs) share the physical wireless substrate operated by the mobile network operators (MNOs). This way, the differentiated offered services coexist on the same infrastructure, sharing its resources and maximizing its utilization. Thanks to this infrastructure-sharing philosophy, the capital expenses and operation expenses of the wireless (radio) access networks can be significantly reduced. From a business point of view, mobile virtual network operators (MVNOs) providing specific services (e.g., VoIP, video calls and over-the-top services) can help MNOs attract more users, while MNOs can produce more revenue by leasing the isolated virtualized networks to them. Moreover, since wireless network virtualization allows the isolation of different network slices, testbeds and research environments can be deployed in real operating networks, providing easier migration to newer products and technologies. Finally, the emerging heterogeneous wireless networks need a convergent and powerful network management mechanism, which can be provided by wireless network virtualization.

Despite the potential vision of wireless network virtualization, it is still at a very early stage of research and a wide range of challenges still remain to be addressed. Unlike wired networks, where bandwidth resource abstraction and isolation can be performed on a hardware basis, radio resource abstraction and isolation is not straightforward. Differently from wired networks, which are reliable, physically isolated from each

other, and have a constant bandwidth, wireless links are less reliable, suffer from interference and have a fluctuating capacity depending on the channel quality.
All these challenges arising from the distinctive properties of the wireless environment make the problem more complicated, and will have to be taken into account when developing a proper solution to meet the expected requirements.

1.2 General Requirements

In general, there are some requirements that need to be satisfied in order to implement wireless network virtualization. These requirements can be classified as basic and additional requirements.

1.2.1 Basic Requirements

a) Coexistence: In wireless network virtualization, physical infrastructures should allow that multiple independent virtual networks coexist on substrate physical networks. Actually, it is clear that the purpose of virtualizing the networks is to make multiple systems to run on the same physical resources. Moreover, these multiple virtual slices have to be able to hold various QoS requirements, topology, services type, and security level and user behavior.
b) Flexibility, manageability and programmability: Freedom in different aspects of networking needs to be provided in wireless network virtualization. However, the flexibility depends on the level at which virtualization is provided, which can range from flow level, sub-channel or time-slot level, or even antennas level. Higher level virtualization may provide less flexibility in resource customization and may reduce the efficacy of isolation. However, the implementation is easier and the complexity of resource multiplexing across slices is reduced. Lower level virtualization leads the reverse. Manageability and programmability are other two basic requirements meaning giving service providers the capabilities to set and modify adaptively virtual network configurations and allocations.

c) Isolation: Ensures that any customization, topology change, misconfiguration and departure of any specific virtual networks will not interfere other coexisting parts. Indeed, virtual slices or virtual networks should transparent to each other, not affecting each other performance. Since many virtual networks should coexist, isolation is the basic issue in virtualization that guarantees fault tolerance, security, and privacy. In addition, in wireless networks, especially cellular networks, any change in one cell may introduce high interference to neighbor cells, and the mobility of end users may create instability of a specific area. Therefore, isolation becomes the most difficult and complicated requirement in wireless networks, specifically compared to the wired counterparts.

1.2.2 Additional Requirements

Heterogeneity, allowing coexistence of different radio access technologies; Scalability, providing capability to support an increasing number of virtual networks; Stability and Convergence, decreasing the effects of errors and misconfigurations in the underlying physical network; Mobility, not only geographical but also between VNs or SPs and Resource Utilization, guaranteeing the efficient use of the available resources are some of the other requirements that should be taken into account when developing a proper solution.

1.3 SoftAir: A Software Defined Networking Architecture for 5G Cellular Systems

This thesis is developed as a part of the SoftAir project [1] from the Broadband Wireless Networking Lab, aiming to propose an SDN based architecture for facing the challenges of the next generation (5G) wireless networks.

The proposed architecture is based on a data plane and a control plane. On the one hand, the data plane is an open, programmable, and virtualizable network forwarding infrastructure, which consists of a software-defined radio access network (SD-RAN)

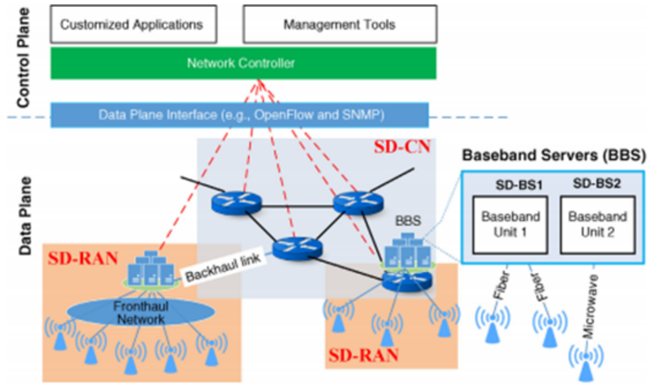


Figure 1: Overall Architecture of SoftAir.

and a software-defined core network (SD-CN). The SD-RAN is formed by a set of software defined base stations (SD-BSs), while the software defined cellular core network (SD-CN) is composed of a collection of SD-switches. In reference to the SD-RAN, a distributed radio access network architecture is proposed, where every SD-BS was split into hardware-only radio heads and software implemented baseband units, not necessarily co-located, where remote radio heads (RRHs), are connected to the baseband units on baseband servers (BBS) through front-haul network (fiber or microwave) using standardized interfaces.

On the other hand, the control plane mainly consists of two components, the network management tools, and the customized applications of the different service providers or virtual network operators. These network management tools are essentially the mobility-aware control traffic balancing, the resource-efficient network virtualization and a distributed and collaborative traffic classifier.

With all these functionalities, SoftAir offers five core properties: (i)programmability, i.e., SDN nodes (e.g., SD-BSs and SD-switches) can be reprogrammed on-the-fly by dynamically associating with different network resources and networking algorithms; (ii)cooperativeness, i.e., SDN nodes can be implemented and aggregated at data centers for joint control and optimization to enhance the global network performance;

(iii)virtualizability, i.e., multiple virtual wireless networks can be created on a single SoftAir, each of which operates under its own independent network protocols with network resources allocated based on demand; (iv) openness, i.e., data plane elements (i.e., BSs and switches), regardless of the underlying forwarding technologies and vendors, have unified data/control interfaces, thus significantly simplifying the data plane monitoring and management; and (v) visibility, i.e., centralized controllers have a global view of the network status collected from BSs and switches, which are essential to enable 5G systems to possess the following promising features: evolvability and adaptiveness, infrastructure-as-a-service, maximal spectral efficiency, convergence of heterogeneous networks, and low carbon footprints.

1.3.1 Network Virtualization in SoftAir

Providing network virtualization in SoftAir is a clear must in order to achieve a scalable, flexible and resilient network architecture. The use of network virtualization capability in Soft Air would enable a wide range of emerging applications, e.g. (1) allowing Multiple Virtual Network Operators to adopt different wireless standards, (2) enabling active RAN sharing, which could allow the reduction in operators equipment investments in low traffic areas, (3) promising virtualization-enabled QoS routing by simultaneously satisfying the strict end-to-end performance requirements (e.g., delay, jitter, and throughput) of different network services that generate fundamentally different traffic flows, and (4) accelerating technology innovation by allocating isolated wireless resources to deploy and test innovative technologies on the operational networks in large-scale real-life scenarios.

To realize network virtualization in SoftAir, as shown in Figure 2, three functions have been proposed:

- The network hypervisor is a high-level resource management framework, which

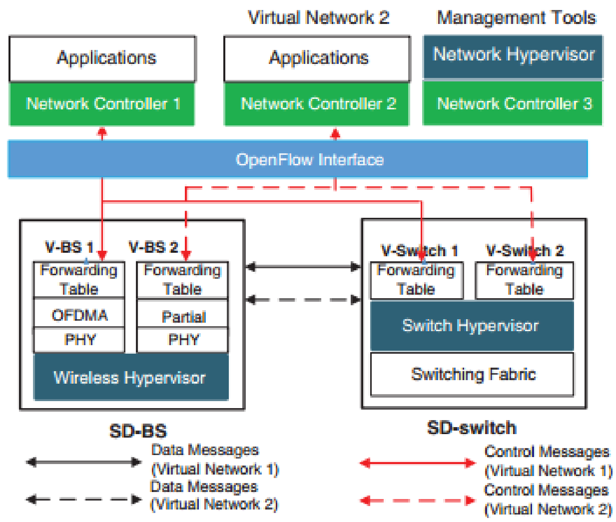


Figure 2: Network Virtualization in SoftAir.

adaptively allocates non-conflicting multi-dimensional network resources to service providers or virtual network operators.

- The wireless hypervisor is a low-level resource scheduler that enforces or executes the resource management polices determined by network hypervisor by employing a variety of wireless resource dimensioning schemes, e.g., OFDMA or wireless scheduling, so that isolation among virtual networks are guaranteed.
- The switch hypervisor focuses on bandwidth management in a single SD-switch, offering link resources assurance for the designated virtual networks.

By utilizing this three level hierarchy with the wireless hypervisor at SD-BSs along with the switch hypervisor at SD-switches, SoftAir will have the capability to enable the end-to-end network virtualization traversing both SD-RAN and SD-CN.
This thesis is focused on the wireless virtualization part of SoftAir, aiming to move towards the design of a throughput-efficiency wireless hypervisor, determining how to distribute non-conflicting network resource blocks among virtual network operators

based on their demands. The objective is to develop utilization-optimal based algorithms to maximize the global resource utilization, while guaranteeing the data rate requirements demanded by each virtual operators.

1.4 Organization of the Thesis

The chapters of the thesis clearly indicate the methodology followed during all the project and are organized as follows. In the second chapter, the design for the network wireless virtualization implementation is introduced, the state of the art is reviewed and a solution for the single-RAT case only considering LTE is developed. Simulations verifying the performance of the presented solution are analyzed and some conclusions and remarks are extracted. In the third chapter, the solution is broadened to a multi-RAT environment where IEEE 802.11 is also modeled and considered. In the last chapter, the different contributions of the thesis are summarized and possible improvements, next steps and future research opportunities are discussed.

CHAPTER II

DESIGNING A WIRELESS HYPERVISOR

Throughout this chapter, the design of network wireless virtualization will be introduced. In the first section, different proposals presented by the research community will be analyzed to extract the key design issues for the wireless hypervisor. Following the observations and the lessons learned from the state of the art, in the second section the problem for the simple scenario of just one BBS with a single RRH and one only radio access technology is formulated. A theoretical study is followed and a mathematical optimization is proposed to achieve an optimal throughput-efficient wireless virtualization. To verify the performance of the algorithm, a simulator is developed and the solution is tested in terms of profit achieved and isolation and compared to other well-known scheduling algorithms such as Maximum Signal to Noise Ratio and Weighted Fair Queuing.

2.1 State of the Art: Analyzing Literature Proposals

Although wireless virtualization has not received all the attention it is entitled to, in the following point the different proposals that the research community has come up will be analyzed, to learn their pros and cons and take them into account when designing our solution.

Before starting the revision of the state of the art, some considerations must be commented. Since wireless virtualization is a young and active area of research, a comprehensive survey of all the recent advances in wireless virtualization technologies is quite difficult to realize. Moreover, the fact that no standards have been developed yet allowing all the research community to walk towards the same path, makes one to find proposals for a very broad range of specific scenarios and situations that do not match the requirements that he is looking for. However, in this section, some of the efforts that have been deployed in this area will try to be summarized, helping us to illustrate the state of the art and the reasons that will bring up to our solution.

There are many different ways of classifying the related work. Special attention will be paid in the depth of virtualization, the metrics defined to optimize, and the techniques employed in each proposal.

2.1.1 Elastic Resource Allocation Enabling Wireless Network Virtualization

An elastic resource allocation algorithm enabling wireless network virtualization is presented in [10] and [9]. The second paper is just an enhancement of the first one taking into account power allocation.

The proposal focuses in resource allocation for WiMAX systems, specifically for the uplink. The allocation is performed in terms of subcarriers. Time Division Duplexing (TDD) is assumed, as it is shown in Figure 3, between the uplink and downlink frames, and the decision, thus, is to select for every time interval, through which subcarriers the flows of the different virtual networks or slices should be able to transmit.

The proposed scheme is dynamic, which means that, at every decision, they consider channel information for allocating the subcarriers to the flows with whom they enjoy the best channel conditions. The scheduling technique that has been designed is the following: Two different kinds of slices are considered, the owner of the physical

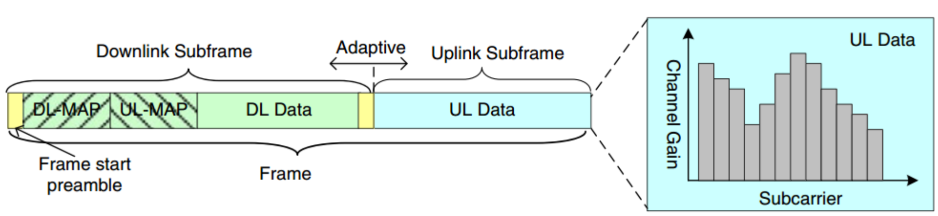


Figure 3: Time Division Duplexing in WiMAX.

resources and the tenants that are renting those resources. The proposed algorithm aims to give the minimum resources possible to the flows of the tenant slices just allowing them to satisfy the needed rate for their QoS requirements, leaving all the rest of the resources, the maximum possible, to the owner slice, so that its throughput can be maximized. Different algorithms are compared to verify the one with best results in solving the optimization and simulations to check both isolation between slices and resource efficiency among owner and tenants slices have also been performed.

Review: On the one hand, this first proposal introduces some interesting ideas. Considering the channel conditions when performing scheduling is always a plus, as it allows the achievement of high levels of efficiency. Current cellular systems offer complex radio resource management mechanisms, giving the possibility of tracking the state of the channel for the different users in a very accurate way. Exploiting these possibilities is a must in the design of efficient scheduling algorithms. Moreover, the idea of working in a flow depth level is the most adequate, allowing the mobile virtual network operators to offer different services with differentiated requirements amongst them and providing isolation at this level of granularity. Finally, trying to satisfy the quality of service requirements is part of the basics of virtualization concept, so it is something we should take into account when developing our proposal.

Although these interesting points will be very relevant for our work, the proposal has also some drawbacks or no-specifications that should be mentioned as well. Working at a subcarrier level adds too much complexity for the computation of decisions and requires a modification of the MAC/PHY layers or an addition of a substrate layer capable of achieving this granularity that the current schedulers do not offer. In particular, the solution is not backwards compatible, making its deployment hazardous for mobile operators. Furthermore, although the solution can work properly in the ideal case in which the resources are unlimited and all the virtual networks can satisfy their requirements, the proposal does not offer any coverage or solution when resources are limited and for example all the tenant slices cannot satisfy their requirements. In that scenario, who has priority? Finally, isolation has not been implemented. If the number of tenants increase, which is a possible situation observing the trend of the future communications towards ultra-densified scenarios, the owner will be starving without being able to transmit a single bit.

2.1.2 Novel LTE Wireless Virtualization Framework

A proposal for wireless virtualization in LTE systems is presented in [18] and [19]. It focuses on the downlink of LTE. LTE downlink is based in OFDM, differently from the uplink that is based in SC-FDMA. This LTE approach does not work at a subcarrier level, but works on a MAC layer level, scheduling Radio Resource Blocks (RBs). A Radio Resource Block (RB) is formed by 12 subcarriers x 7 OFDMA symbols and is the smallest unit that the LTE MAC scheduler can allocate. The problem then is to schedule RBs among different mobile virtual network operators (or slices) in a fair way. The proposed technique is based on contracts. Four different

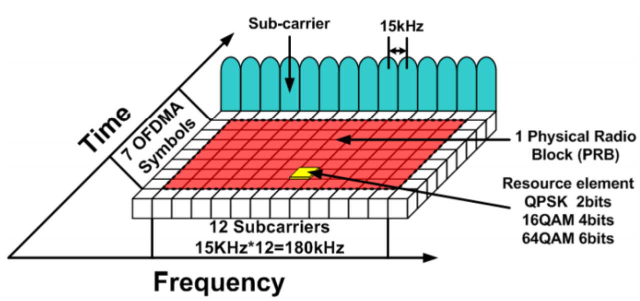


Figure 3. LTE Downlink Physical Resources Structure

Figure 4: LTE Resource Block.

kinds of contracts are defined: Fixed Guarantees, Dynamic Guarantees, Best Effort with minimum Guarantees and Best Effort without Guarantees. Each of them have assigned different priorities and specifications and the division of RBs is based on these requirements. Between slices with same priority, the scheduling is performed according to bandwidth requirement estimations. The proposed procedure is the following:

1. Each operator has to estimate the BW that it needs at frequent time intervals:

$$E(n) = \alpha \cdot E(n-1) + (1-\alpha) \cdot RBs\_TTI(n) \tag{1}$$

2. Fairness weight is computed:

$$FF_i = E(n)_i / Total\_Est \tag{2}$$

3. Allocation is computed:

$$RBs\_Alloc_i = int(FF_i \times Left\_RBs) \tag{3}$$

As one can observe, to distinguish between virtual networks with the same contract, the proposed algorithm is a modification of the well-known Weighted Fair Queuing or GPS for resource blocks, giving every virtual network a proportion of resources according to what they need.

Review: The best insight of this proposal is that it works at a RB level, without needing this way any extra modification layer, being backwards compatible with current specifications. Another good point is that they try to consider fairness between virtual networks with the same condition, trying to provide this way isolation between them. However this has two cons associated. The first one is that, with the fairness mechanism proposed, there is no guarantees that the requirements of the slices will be fulfilled. In that case, only a proportional amount of data of the actual needed would be transmit without actually satisfying the QoS requirements, data that for these reason could probably be discarded when arriving to the user for being incomplete, and wasting useless power of the BS. Moreover, with the modified WFQ that they present, there is no isolation between slices. It happens the same as with the previous proposal, if one virtual network requests a big amount of resources acting selfishly (because at the end, the estimations will be done and given by the mobile virtual network operators), they will receive a proportion of the resources, letting the other virtual networks without capability to transmit. Probably the biggest drawback of the proposal is the fact of not considering channel information, basing the scheduling in fixed contracts instead of current channel conditions, making the proposal absolutely inefficient in terms of resource utilization.

2.1.3 NVS: A Virtualization Substrate for WiMAX Networks

NVS is the more complete solution at the moment. A framework that allows flow level virtualization for WiMAX networks is proposed in [8]. Same as LTE Novel, the approach is to modify the existing scheduling algorithms to isolate and differentiate groups of users. In WiMAX the BS is the one that develops the Radio Resource Management (as well as QoS guarantees and Handoff Triggering), and uses the OFDMA frame structure to schedule downlink and uplink transmissions between clients. Furthermore, a traffic classifier splits packets in flows, according to their different QoS. NVS takes profit of these WiMAX properties to consider flows in groups (slices) and schedule resources amongst them.

The scheduling problem is divided in two steps: slice scheduling and flow scheduling. Slice scheduling is the process of selecting which slice has to transmit at every moment. This first decision is done based on utility maximization. Every slice can request a certain amount of resources or bandwidth, noting that with requesting resources, the real bandwidth allocated cannot be expected as depending on channel conditions fixed amount of blocks can provide different bandwidths. Here bandwidth is referred to the throughput rate achievable by a packet. Once all the slice requirements are set, at every moment the slice that maximizes the global utilization will be selected. The second step of the process is not further specified. The selected slice can customize the flow scheduling, selecting this way which flow should transmit in the first position. Three different kinds of flow scheduling are proposed according to the slice objectives, choosing from a general schedulers list, a weight distribution as a function of the average rate already achieved, or letting the slice perform the flow scheduling itself. Finally, when both decisions have been taken, the frame scheduler will be invoked to perform the allocation of the resources to the resulting packet.

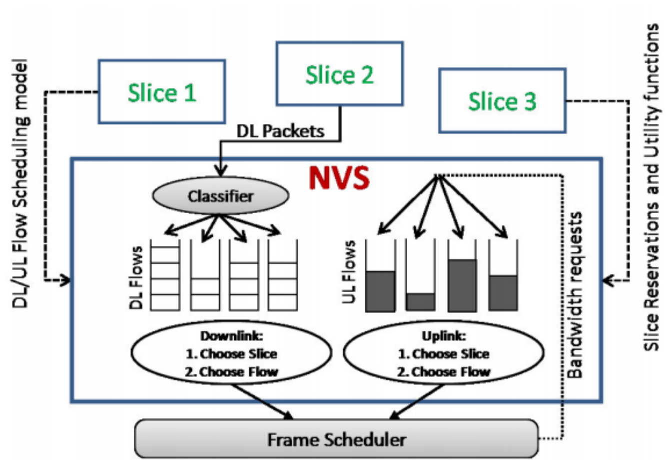


Figure 5: NVS Structure.

Review: The idea of maximizing the global utility is great when moving towards a throughput efficient solution. And the willingness of meeting the rate requirements of the different virtual networks is also a good point as we have discussed previously. Working at RB level, as we also mentioned before is a good point, and allows the usage of already deployed systems without need of updating them. However, there are some drawbacks that should be remarked. The fact of being the virtual networks the ones requesting a specific requirement instead of their flows weaknesses the flow-granularity that we were looking for. And at the same time, it adds complexity for the virtual networks to determine how many resources they need for all their flows, and to decide how to distribute these resources among them. Sometimes, trying to give more flexibility what we do is just increase complexity for third parties and this is a good example. Making virtual networks decide how many RB they will need makes the algorithm easier, but is it a practical solution?

2.1.4 Conclusions

After reviewing the most illustrative approaches that have been proposed some conclusions can be extracted, to be considered for the design of our wireless virtualization solution. As it has been shown throughout the state of the art review, virtualization can be done at different levels: flow level (assigning RBs to flow packets), sub-carrier or time slot level, or even at the lowest level of hardware components (antennas and signal processors). Virtualization at higher levels leads to a better multiplexing of resources across slices (and hence increased utilization with fluctuating traffic), and simplicity of implementation, but at the same time can reduce the efficiency of isolation and the flexibility of resource customization. Whereas, virtualization at lower level leads to a reverse effect. Taking into account that MVNOs often require only flow-level isolation, caring only about flow level or application level QoS, that large number of research efforts on developing innovating flow level techniques for which lower level virtualization is not required and that too much flexibility would expose unnecessary complexity to slice owners, seems accurate to believe that flow-level virtualization is appropriate. Furthermore, this will provide flexibility when broadening the range of radio access techniques in the next chapter, where some important technologies such as WiFi do not necessarily use OFDMA, leaving the sub-carrier level choice offside. In the following lines some of the decisions extracted for the design of our proposal are summarized:

	Scenario	**Depth Level**	**Resource**
Elastics	WiMAX Uplink	Flows and Slices	OFDMA Subcarriers
Novel LTE	LTE Downlink	Slices = MVNOs	RBs
NVS	WiMAX Uplink+Downlink	Slice and Flow	MAC-Frame

Table 1: Comparison State of the Art.

- The first objective commented at the beginning of the thesis was providing isolation. Isolation among slices, allowing them to deploy their applications in a way that the performance of their neighbor ones does not affect their own one. Isolation in terms of interference can be achieved by working with orthogonal inter-cell mechanisms such as OFDMA and controlling intra-cell power. But a mechanism for resource sharing isolation needs to be developed, allowing a virtual network to increase its rate or its QoS requirement, or even its number of users without affecting the other virtual networks when possible.

- Providing backwards compatibility, which means developing a solution that do not need modification of the MAC/PHY layers is another interesting challenge. The scheduling has to be performed according to the existing specifications. Schedulers in current cellular systems work with a granularity of a certain transmission time interval (TTI) i.e. 1ms for LTE and a resource block (RB) in time and frequency respectively. Working with the same granularity is a good option to provide the aimed backwards compatibility.

- One of the strongest requirements is computational complexity. As it has been remarked during all this chapter, it is highly important to take scheduling decisions in a very quick way. The use of Software Defined Networking architectures allows the usage of high power computation by performing the scheduling algorithms in the network controllers, located in specific data centers. However, finding the best allocation decision through complex optimizations or exhaustive research over all the possible combinations could be too expensive. For this reason, fast algorithms will have to be developed to make the scheduling computation as easy as possible.

- A channel-aware solution needs to be adapted if a throughput efficient solution is aimed. Giving high priorities to flows with better channel conditions is the only way to achieve high spectral efficiencies. Taking into account the radio resource management (RRM) techniques that current cellular systems contain (i.e. CQI reporting) when performing the scheduling is a must for any new generation scheduler.

- Satisfaction of the QoS requirements of the different slices flows but balancing it with fairness among slices, flows and users and trying to maximize throughput is a good focus for design.

Following these key design points obtained from the state of the art analysis, a proper solution for efficient wireless virtualization can be developed. Our proposal is based on three main concepts:

Profit: Comparing the throughput needed by each virtual-network flow to satisfy its QoS requirements with the throughput offered by the channel at each moment allows to perform the best allocation at every scheduling decision and gives high levels of efficiency to the algorithm by selecting the flow that takes the most profit from the channel conditions at each moment. Moreover, detecting if the delay requirements of the service providers can be fulfilled avoids useless data transmissions and retransmissions in case of not accomplishing them. It is based on a supposition of full radio resource management (RRM) techniques availability.

Fairness: Responsible for the isolation control at three levels: slice, flow and user, and providing the enough flexibility to the MNO to decide which policy of differentiation between them does it want to assign. With fairness index, on the one hand, the amount with which a specific slice operation affects the other slices can be controlled. Moreover, if a user has bad channel conditions making it obtain bad levels of profit, it allows the system to give him also chances to transmit, avoiding starvation. Finally, if based on contracts or any other reason, more priority is desired to be given to a specific virtual network or even to a specific service of this network, fairness indexes give the possibility to do it.

Power: Controlling the transmission capabilities of the system together with the second QoS requirement, the minimum signal to noise ratio in reception, allows the optimal distribution of power between the parties improving the global performance of the network and achieving a second level of efficiency.

By exploiting these three points a practical but efficient and flexible solution is proposed. The reader may feel that there are multiple possible combinations of these three factors that can result in a good scheduling algorithm. And that is right, we could have decided to make a triple minimization considering the three variables (power, fairness and profit). However, we have decided to give more importance to the latter one formulating the problem as a single optimization with the fairness and power acting just as limiting bounds as we believe that the proposed methodology is the one conceptually clearer and the one allowing less computational complexity respect multiple optimization cases. Actually, the idea of the project is assenting the basis and key issues in the design of the next generation cellular systems wireless virtualization, and obviously the possibilities of methods and algorithms for doing it is as broad as the imagination allows. For this reason, at the end of the document some conclusions from our design will be exposed and interesting research future steps will be discussed.

2.2 *Problem Formulation*

Suppose the simple scenario shown in Figure 6 where just one base station with a base band server (BBS) and a single remote radio head (RRH) is considered. As 5G is still at an early stage of research, with a lot of open possibilities but without any standard nor specification having been developed yet, drawing a common path that we could strictly follow, and specifically in terms of scheduling, where any change respect the last advancement in LTE technologies has been agreed, it will be supposed that the base station supports LTE. At the end, what is clear is that, differently from the jumping between previous generations, the step from 4G to 5G will not be that substantial. However, we will try to provide the enough flexibility in our solution to easily incorporate possible enhancements that 5G could finally include.

Now assume that, to this base station, on the one hand, different virtual networks of different mobile virtual network operators (MVNOs), not necessarily mapped one to one (i.e. one MVNO can own one or several slices), are allocated. On the other hand, some users are connected to this BS, all of them spread around the cell occupying different positions, and thus, observing different channel conditions. These users are subscribed to some of the services that the virtual networks of the BS are providing. A single virtual network can offer a highly diverse range of services, all of them demanding different QoS requirements. This brings up the concept of flow. Every association in which a user is receiving services from a virtual network is called flow. A flow, thus, is characterized for being a group of data of a specific virtual network service, designated to a specific user. Each flow has specific but not necessarily unique QoS requirements. One can think on the "Snapchat" virtual network for example, which could have one flow for static image sending service and another flow for real time video calls. Each of both services would share the same virtual network but they would have different requirements. Video real time would probably require higher throughput allowing more packet losses, while static image would require less

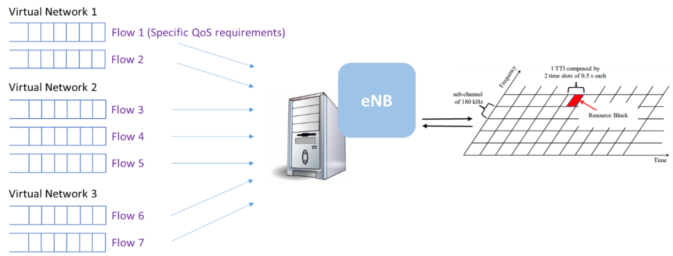


Figure 6: Wireless Virtualization Scenario.

packet losses but could allow a lower rate. This way, "Snapchat" virtual network would have two different flows for every subscribed user, one for image service with specific imageQoS requirements and one for video service with specific videoQoS requirements.

To deliver the data of the different flows to their corresponding users, the BS needs to allocate the physical resources to them. In LTE resource allocation is performed in terms of resource blocks (RB) and the decisions are taken every time intervals of 1ms, called transmission time intervals (TTI). A radio block is formed by 12 sub-carriers, allocating 7 OFDMA symbols each one. In the time domain, one RB occupies a slot of 0.5 ms. At each TTI, the BS (or eNB if preferred by now) has to assign one or several resource blocks to each of the different flows [5]. Note that we are only focusing on downlink.

The objective, thus, is to find the best way to allocate the RBs to the different flows of the base station in an efficient way (i.e. taking the maximum profit of the available resources) and strictly satisfying the QoS requirements specified by the flows (i.e. if one flow needs 2 RB to satisfy its requirements but only 1RB is available, the RB will not be assigned to that flow, giving this way the chance to another flow, that in this case only needs one RB, to satisfy its requirements).

To take the decision, the BS or eNB has different information available. Every flow of every base station has two kinds of QoS requirements. The one that we will treat first is the called QoSDelayBudget, which refers to the maximum time that the flow data can stand before arriving to the UE. In other words, the data received at the base station of a specific flow, will have to be delivered to the user attached to that flow in less than the QoSDelayBudget time. This way:

$$Data\_required_f = Arrival\_rate_f \cdot TTI \quad (+buffer) \tag{4}$$

$$Throughput\_required_f = \frac{Data\_required_f}{QoSDelay_{min,f}} \tag{5}$$

The satisfaction of this requirement depends basically on the number of resources (RBs in our case) that are allocated to the flow, and the amount of information that can be carried in each of those RBs.

Considering full RRM techniques availability, it is assumed that channel quality information (CQI) reports from all the users are received periodically by the base station, knowing this way how much information can be transmitted to them according to their channel sate at every moment. This information is used for the eNodeB to select the most appropriate modulation and coding scheme according to the channel conditions at that moment. The modulation is fixed by the CQI, but the code rate depends on the Transport Block size selected by the eNodeB, which will try to find the best combination to fulfill the code efficiency stated by the CQI report. The code rate is defined as the ratio between the transport block size and the total number of physical layer bits per sub-frame that are available for transmission of that transport block. For the project purposes, we will just suppose that the code efficiency achieved with the MCS of this selection is directly the one defined by the CQI report as described in the table below. Once a RB is assigned to transmit data of one of the users flows, depending on the user channel conditions, a specific amount of data will be able to be transmitted. Thus, considering that all OFDMA symbols are used for data:

$$Data\_offered_{r,f} = 7 \cdot 12 \cdot Modulation\_size_{u[f]} \cdot Code\_efficiency_{u[f]} \tag{6}$$

CQI index	modulation	code rate x 1024	efficiency
0	out of range		
1	QPSK	78	0.1523
2	QPSK	120	0.2344
3	QPSK	193	0.3770
4	QPSK	308	0.6016
5	QPSK	449	0.8770
6	QPSK	602	1.1758
7	16QAM	378	1.4766
8	16QAM	490	1.9141
9	16QAM	616	2.4063
10	64QAM	466	2.7305
11	64QAM	567	3.3223
12	64QAM	666	3.9023
13	64QAM	772	4.5234
14	64QAM	873	5.1152
15	64QAM	948	5.5547

Figure 7: Channel Quality Indicator LTE.

is the data that the flow *f*, destined to the user u, can transmit if it is assigned the RB r, according to the modulation and code efficiency selected. Note that the data offered by the channel does not depend on the flow it is assigned, only depends on the user the flow is attached to. The amount of data that the flow f from the user u can transmit from the whole LTE RB bandwidth in one scheduling decision is:

$$Data\_offered_f = \sum_{r=1}^{RB} i_{r,f} \cdot Data\_offered_{r,f} \tag{7}$$

Where $i_{r,f} = 1$ when the RB r is assigned to the flow f and 0 otherwise.

$$\sum_{f \in F} \sum_{r \in RB} i_{r,f} \leq RB\_BW \tag{8}$$

This way, the profit that the flow is taking from the offered resources measures the difference between the throughput needed to satisfy the QoS requirements and the one offered by the channel. It can be defined as:

$$Profit_f\,(\%) = \begin{cases} \frac{Throughput\_required_f}{Throughput\_offered_f} \cdot 100 \\ \\ 0 \;\; if \sum_{r=1}^{RB} i_{r,f} = 0 \end{cases} \tag{9}$$

Our main objective is to maximize the profit of all the flows, allowing this way the maximum satisfaction of their requirements. The optimization solving the problem could be written as:

$$\max_{i_{r,f}} \frac{1}{F} \sum_{f \in F} Profit_f \tag{10}$$

$$s.t. \quad Profit_f \leq 100 \quad \forall\, f \tag{11}$$

But that is not enough. We can find a situation in which users very close to the BS, which have optimum channel conditions, are subscribed to high throughput demanding services, creating a combination in which the profit of the channel is maximized. In this scenario, only the flows of these users with that specific characteristics would be the ones receiving the maximum amount of resources all the time, leaving other users with high requirements and bad channel conditions without possibility to transmit. At the beginning of the thesis we said that one of the important points to take into account when designing virtualization techniques was isolation. And isolation meant that the performance of one virtual network or flow, could not affect the performance of the other virtual networks. Particularly, any virtual network could increase the number of services offered, the data rate requested or even the number of users without harming the performance of the neighbor virtual networks. Ideally, the performance of every virtual network should be the same regardless of the existence and performance of other virtual networks sharing its resources. Obviously this is not possible as when having only one virtual network we can allocate all the resources to it but if we have more virtual networks to distribute resources amongst, less resources can be allocated to each of them. However, the idea is that some kind of control had to be included in our design, to allow all the flows of all the virtual networks to have at least some chances to transmit.

To do it we have developed the concept of fairness. Fairness is basically a historical record of the profit obtained by each flow, stored in the base station since the allocation of the virtual network in the base station took place and updated at every decision, that allows the base station to have an idea of how much utilization of the channel resources has the flow been achieving. This way, if in the next decision a specific flow is selected to transmit, the new profit will be averaged with the historical value saved, updating dynamically the parameter, so that if the new profit is high, the value will increase and if the profit is low the value will decrease. And if the specific flow is not selected, the historical value will be averaged with a zero, making decrease the fairness parameter. The idea is that if the fairness of a flow is going too down, more priority to that flow can be given when performing the next decision. In the same way, if the fairness of a flow is going too high maybe it is convenient to give priority to the other flows even though the profit that they can obtain is not that big, controlling this way the system isolation, giving equity and opportunities between all the flows.

For all the selected flows:

$$Fairness_f\left(d\right) = \frac{Fairness_f\left(d-1\right) \cdot D_{d-1} \quad + \quad Profit_f\left(d\right)}{D_d} \tag{12}$$

And if they have not been chosen:

$$Fairness_X\left(d\right) = \frac{Fairness_X\left(d-1\right) \cdot D_{X\,d-1}}{D_{X\,d}} \tag{13}$$

Where d is the decision index and D stands for the number of decisions taken at the *subindex* instant, since the slice, user or flow were connected to the BS. We just need to add a new constraint in the optimization:

$$\varphi_1 \leq Fairness_f \leq \varphi_2 \quad \forall\ f \tag{14}$$

This idea of fairness, to control the impact of every flow in the network is very useful when sharing resources. It is very common to have situations in which different parties impact on the network needs to be controlled, not only at a flow level, but also at a virtual network level. An example of that could be the differentiation of contracts for example. We have seen in the state of the art review that a lot of proposals consider scenarios in which the different virtual networks are treated differently based on the possible contracts with the MNO. We may also need to provide equity between users of a virtual network in terms of transmission opportunities when allocating the resources. All these scenarios depend on the policies of the operator and the degree of customization when virtualizing can be as big as desired. As we wanted to provide a flexible solution and we have seen that by exploiting the fairness concept all these regulations can be implemented, we decided to broaden the idea of fairness at three levels, flow (as we had), but also virtual network (or slice) and user. The followed methodology is exactly the same as mentioned for the flow case, but now the historical parameter of resource utilization is computed for the virtual networks and the users, with their corresponding flows information. F_u and F_s are flows for user u and slice s respectively.

$$i_f = 0 \quad when \quad \sum_{r=1}^{RB} i_{r,f} = 0 \ , \quad i_f = 1 \ otherwise \tag{15}$$

$$Profit_u = \frac{1}{F_u} \sum_{f \in F_u} i_f \cdot Profit_f \tag{16}$$

$$Profit_s = \frac{1}{F_s} \sum_{f \in F_s} i_f \cdot Profit_f \tag{17}$$

For all users:

$$Fairness_u\,(d) = \frac{Fairness_u\,(d-1) \cdot D_{d-1} \quad + \quad Profit_u\,(d)}{D_d} \tag{18}$$

For all slices:

$$Fairness_s\,(d) = \frac{Fairness_s\,(d-1) \cdot D_{d-1} \quad + \quad Profit_s\,(d)}{D_d} \tag{19}$$

Now that the two first points of our proposal, profit and fairness, have been analyzed, it is time to define the last one, power. In the beginning of the section we said that the different flows of the base station had two quality of service requirements that needed to be satisfied. The first one was the QoSDelayBudget, giving the maximum time that the data could support before being handed in to the user. The second requirement that the flows of our base station have is the minimum signal to noise ratio (SNRmin) that has to be achieved in the reception end. As explained with the example of "Snapchat" virtual network, depending on the type of service that the virtual networks are offering, sometimes is not only necessary to maintain a specific throughput but also a certain quality in reception has to be achieved. In the case of a static imaging service, the delay may not be very relevant (the user can receive a picture 1 second late without this supposing any trouble in his experience), but the received quality may be a must (especially when dealing with high quality pictures). To achieve a certain quality in reception, for a specific throughput already established, the base station needs to transmit the signal at a specific power. This power not only depends on the quality wanted to achieve, but also on the channel conditions, the gain of the antennas and the path losses.

$$Power\_required_f = \frac{Throughput\_required_f \cdot \left(\frac{Eb}{No}\right)_{min,f} \cdot \left(\frac{P_N}{B}\right) \cdot MI \cdot MF \cdot L\left(d_{u[f]}\right)}{G_T \cdot G_R} \tag{20}$$

where G_T, G_R, MI, MF, P_N, B and L are the transmission and reception antenna gains, the interference and fading margins, the noise power, the bandwidth and the propagation losses of the channel, respectively.

However, following the stated assumption of full RRM mechanisms availability, including link adaptation, the values of the channel instantaneous condition can be obtained the following way. CQI gives the maximum modulation and code rate that can be used according to the channel conditions, so that a block error rate of 0.1 is maintained at the UE [14]. If we analyze how the CQI is the operation performed for the CQI computation, it can be mathematically described as:

$$Rb_{\max} = r \cdot m \cdot B = \frac{P_{T\_\mathrm{CQI}} \cdot G_T \cdot G_R}{\left(\frac{P_N}{B}\right) \cdot \left(\frac{Eb}{No}\right)_{\min\_CQI} \cdot MI \cdot MF \cdot L\left(d\left[u\right]\right)} \tag{21}$$

In which r and m are found according to a $P_T = \frac{P_{eNB}}{RB}$, $B = 12$ sub-carriers $\cdot$ 15000 Hz, and $\left(\frac{Eb}{No}\right)_{\min\_CQI}$ so that $BLER < 10\%$ for the current MCS.

This way:

$$\frac{G_T \cdot G_R}{\left(\frac{P_N}{B}\right) \cdot MI \cdot MF \cdot L\left(d\left[u\right]\right)} = \frac{r \cdot m \cdot B \cdot \left(\frac{Eb}{No}\right)_{\min\_CQI}}{P_{T\_CQI}} \tag{22}$$

so the power required to achieve a more strict Eb/No requirement with the same throughput, code rate and modulation can be given by:

$$Power\_required_f = \frac{Throughput\_required_f \cdot \left(\frac{Eb}{No}\right)_{min,f} \cdot P_{T\_CQI}}{r \cdot m \cdot B \cdot \left(\frac{Eb}{No}\right)_{\min\_CQI}} \tag{23}$$

where all of them are known. Power allocation can suppose and important improvement if SNR requirements are desired to be satisfied.

$$Total\_power = \sum_{f \in F} i_f \cdot Power\_required_f \tag{24}$$

Although the initial objective of the thesis only aimed to satisfy data rate requirements, verifying if the power available in the base station for a specific radio access technology is enough for transmitting the amount of data required with the energy needed to satisfy their demanded quality, will be just reflected as another constrain bounding the research results in the proposed optimization.

Taking into account all the parts that have been discussed, the problem can be formulated as:

$$\max_{i_{r,f}} \sum_{f \in F} \quad Profit_f \tag{25}$$

$$s.t. \quad Profit_f \leq 100 \quad \forall\, f \tag{26}$$

$$\varphi_1 \leq Fairness_{u,f,s} \leq \varphi_2 \quad \forall\, u, f,\ s \tag{27}$$

$$Total\_power \leq P_{LTE\ RAT} \tag{28}$$

2.3 Creating a Simulator from Scratch: Object Oriented Programming for Virtualization

Once the theoretical study has been proposed, the next objective was to verify the performance of the developed solution in a different variety of scenarios, being able to analyze the efficiency levels achieved, the effects of fairness control techniques and most important, to compare the results with the ones obtained with the other typical scheduling algorithms broadly used in literature such as Weighted Fair Queuing and Maximum SNR.

To do it, any of the simulators developed by the research community such as LTE Vienna Simulators [11] or NS2 or even private software such as LTE Mac Lab or OPNET could have been used. All of them try to simulate real LTE systems so that every researcher can deploy his solutions and verify their performance in a realistic an accurate environment. However, letting a side the fact that the understanding of these huge and complex simulators requires huge amount of time and experience, these proposed simulators do not offer possibilities for the creation of virtual networks and they only rely in a single radio access technology (for example LTE), making impossible to develop solutions in which multi-RAT is considered. For these reason, knowing that our idea is to extend the algorithm for multiple technologies, a simulator from scratch has been developed. Developing the simulator from scratch allows us to have complete control of every incident occurring in the system and permits us to focus only on the part of the technology that we are interested (i.e. focusing only in the scheduling mechanisms and avoiding problems of configurations and issues that may arise in the physical layer part of the program for example in case we do not need it). We developed the simulator completely in Java. Java is an object oriented programming language, allowing the creation of different classes and objects and the inter relation between them. The use of object oriented methodology facilitates both the coding and the understanding of the scripts as it adapts perfectly to the reality, being able to create environments of virtual networks, flows, slices, radio blocks and users same way they operate in the real systems.

The only drawback of implementing the simulator in Java is that, differently from other languages widely used such as MATLAB (that were not considered for giving too little inter-class flexibility) that can plot graphs and results with just writing a simple script, Java does not offer a visual interface to generate plots and charts of the obtained results. To solve this problem, a library called Apache POI [12], that basically allows the exportation of desired results from the simulations to a Microsoft Excel sheet, was installed, being able this way to create all the complex plots, graphs and charts needed under the Excel capabilities.

2.3.1 Simulator Overview

In the UML class diagram from Figure 8, the different classes, attributes and methods of the developed simulator, as well as their relations, can be observed.

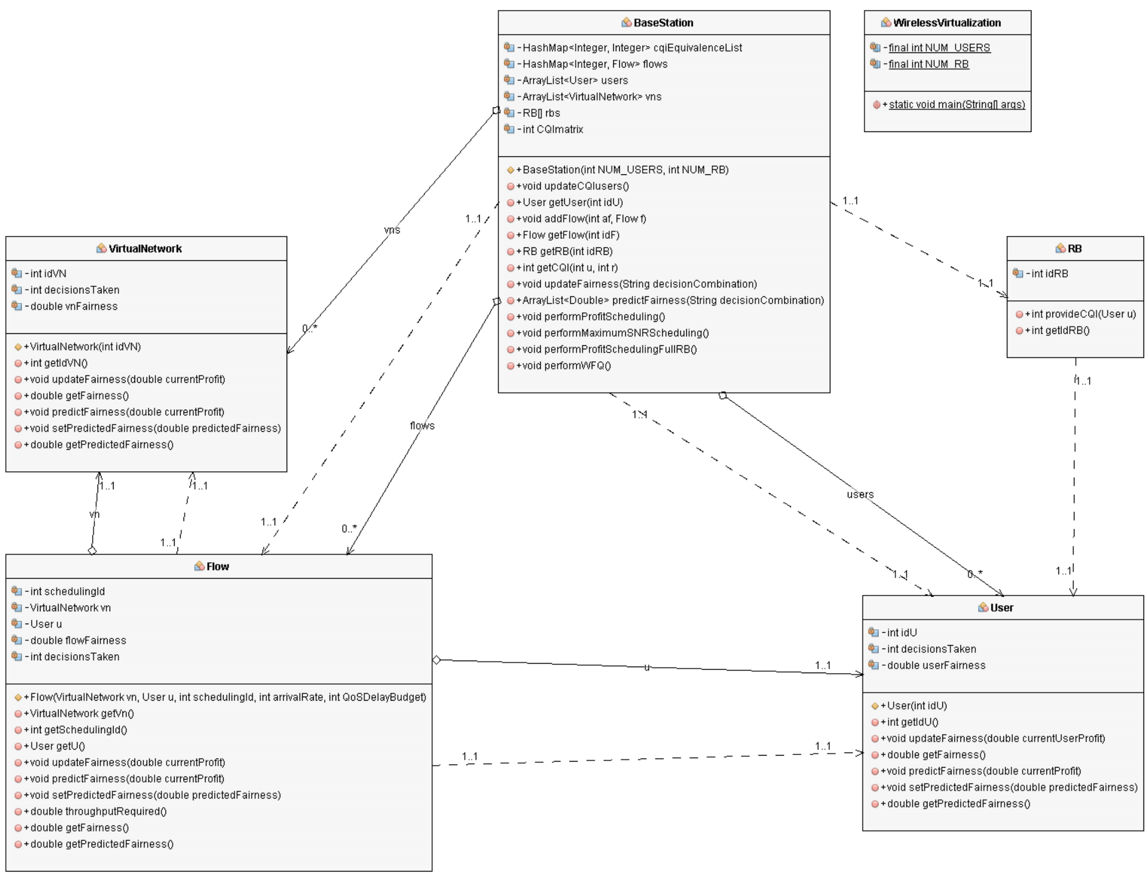


Figure 8: Simulator UML Class diagram.

All the program moves around a central class, *BaseStation*, that contains all the current flows allocated to it, and a list of all the users and virtual networks connected. On the other hand it also contains the different RBs available that need to be allocated and LTE technical information such as the CQI information to map the feedback reports received to their equivalent rates and modulations. The most important methods from the *BaseStation* class are:

updateCQIusers(): Gets periodically all the channel state reports from the users.

updateFairness(decisionCombination): Sends the combination of flows allocated at each decision to all users, virtual networks and flows so that the fairness parameter can be updated accordingly.

predictFairness(combination): Returns the fairness parameters of the virtual networks, flows and users that would be obtained in case the combination passed by value was selected as a final decision.

The second most important class that also deserves to be briefly explained is *Flow*. Every *Flow* has a unique id that allows the BaseStation to compute all the possible combinations and also to get specific information from them. Also a user associated, that will be responsible for the CQI reports and power computation, and the virtual network it forms part of. Leaving a side helping methods such as *computeFairness*, *getters* and *setters*, the *Flow* class contains the arrivalRate and QoSdelaybudget, both of them set by the constructor, that allow the computation of the throughput required through the method specified with the same name.

The rest of classes, *User* and *VirtualNetwork* can be adaptively allocated to the *BaseStation*. Their relations are as follows: Every *User* can be receiving services from different *Flows*, but every Flow is attached to a single user. In the same way, every *VirtualNetwork* can contain different *Flows* but every *Flow* is just part of a single *VirtualNetwork*.

The full simulator is focused on a specific method called *performProfitScheduling*, which, as the name makes explicit, performs the scheduling of the different resources available to the different base station flows. Basically, it selects the amount of resources that should be allocated to every flow at every decision. To do it, it solves the optimization proposed in Section 2.2 by using an exhaustive research methodology, computing first all the possible combinations of radio blocks among the different flows and selecting the best combination for every decision.

performProfitScheduling: Exhaustive research implementation

Output: Optimum resource allocation combination

Compute all possible combinations of RBs among flows

For each combination:

    For all the selected flows in the combination:

        - Obtain throughput required by the flow from the data received and the delay requirement

        - Obtain the CQI of the flow associated user

        - Compute throughput offered to the flow user with the current combination of RBs

        - Compute flow profit

        If flow throughput requirement not satisfied:

            Discard combination

        Else

            Update profit of current combination

    End

    If (profitCurrentCombination > profitDecisionCombination)

    & Fairness and Power constraints satisfied:

        decisionCombination = currentCombination

End

Perform decision (decisionCombination)

Update Fairness (decisionCombination)

2.3.2 Simulation Results

The objective of this section is to verify the performance of the designed algorithm in different scenarios with different traffic conditions to analyze the results in terms of profit levels obtained, throughput efficiency, comparing our solution with other well-known scheduling algorithms and fairness control. The simulator flexibility allows the creation of diverse flow, virtual network, user and requirement configurations to test the optimality of the scheduling decisions under a wide range of conditions.

2.3.2.1 Profit Simulations

The first set of simulations aim to analyze the behavior of the proposed solution in terms of profit. In order to achieve a throughput-efficient solution, the profit metric has been defined as the relation between the throughput required by each flow to satisfy the delay QoS requirements and the throughput that can be achieved by their user depending on the instant channel conditions. Making a comparison between these two parameters, the best allocation of channel resources to flows can be found. Specifically, the best allocation is found in the combination of flows that maximize the profit of their respective channels (i.e. the throughput that thee flows need to transmit is closer to the one they are being offered, but below).

Three different scenarios have been compared, with different virtual networks allocated to the base station.

In the first scenario, three virtual networks are created, the first one offering services of high definition video and online real time gaming, the second one providing web services and the third one offering VoIP and VideoCall services. In Table 2, the specific QoS requirements in terms of delay and packet loss of each service, and the arrival rate specifications have been summarized. All the values have been selected following the LTE Quality Class Information specified by 3GPP.

This first scenario has three subscribed users, connected to the base station to receive the services of the allocated virtual networks, generating the flows from Table 3.

Moreover, users channel conditions have been set completely random, varying within a range of values from CQI=1 to CQI=15, all of them with the same probability and

	E2E Delay	**BS-UE Delay**	**Packet Loss**	**Data Rate**
VN0				
HDTV	300 ms	280 ms	10e-6	30 Mbps
Real Time Gaming	50 ms	20 ms	10e-3	10 Mbps
VN1				
File Transfer	300 ms	280 ms	10e-6	6 Mbps
Web Browsing	200 ms	180 ms	10e-6	5 Mbps
VN2				
VoIP	100 ms	80 ms	10e-2	80 Kbps
VideoCall	150 ms	130 ms	10e-3	1 Mbps

Table 2: First Scenario Specifications

Flow id	**User id**	**VN id**	**Service**	**Arrival Rate**	**Delay Budget**
1	1	1	HDTV	30000 kbps	280 ms
2	1	2	File Transfer	6000 kbps	280 ms
3	1	3	VoIP	80 kbps	80 ms
4	2	1	HDTV	30000 kbps	280 ms
5	2	1	File Transfer	6000 kbps	280 ms
6	2	2	Web Browsing	5000 kbps	180 ms
7	3	1	RT Gaming	10000 kbps	30 ms
8	3	3	VoIP	80 kbps	80 ms
9	3	3	VideoCall	1000 kbps	130 ms

Table 3: First Scenario Configuration.

with a feedback periodicity of one decision. The base station allocated flows have been set as static (i.e. no new flows were created or removed during the scheduling process), and no fairness nor power control have been applied. Finally, all the data not selected for scheduling has been assumed lost, which means that no buffering queue has been considered.

For this first scenario, three different situations have been taken into account, the first one being characterized for having 1RB available at each decision to allocate, the second one having 2RBs to allocate and the third one having 4RBs to allocate. The obtained results can be shown in the graphs below.

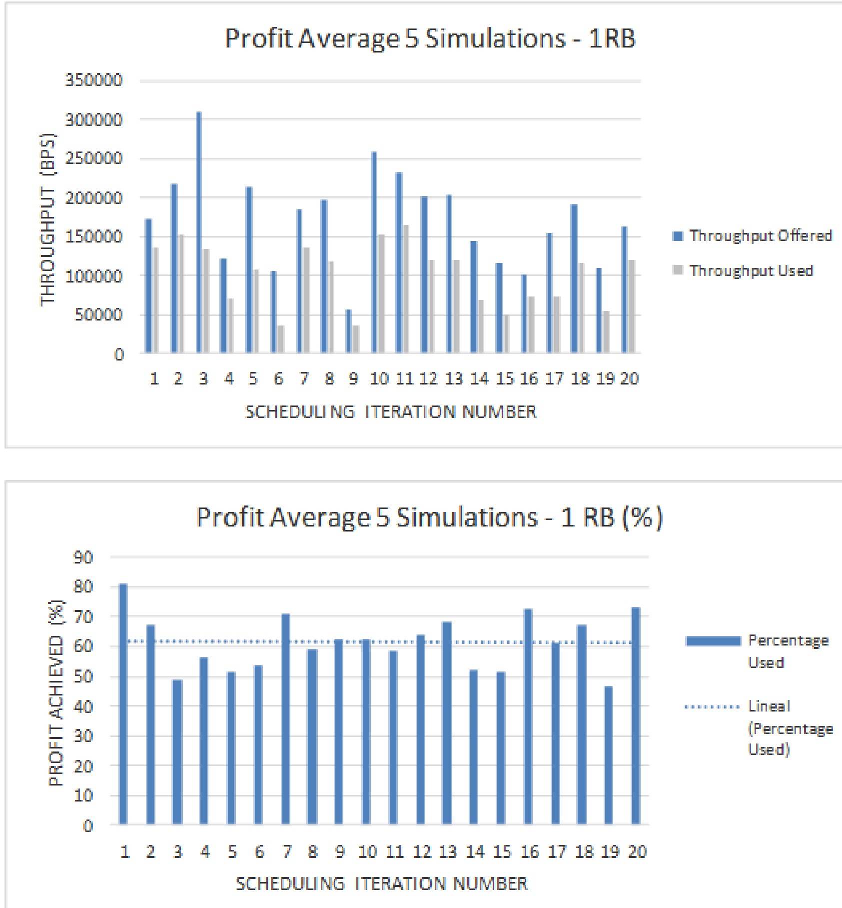


Figure 9: Profit Simulation First Scenario 1RB.

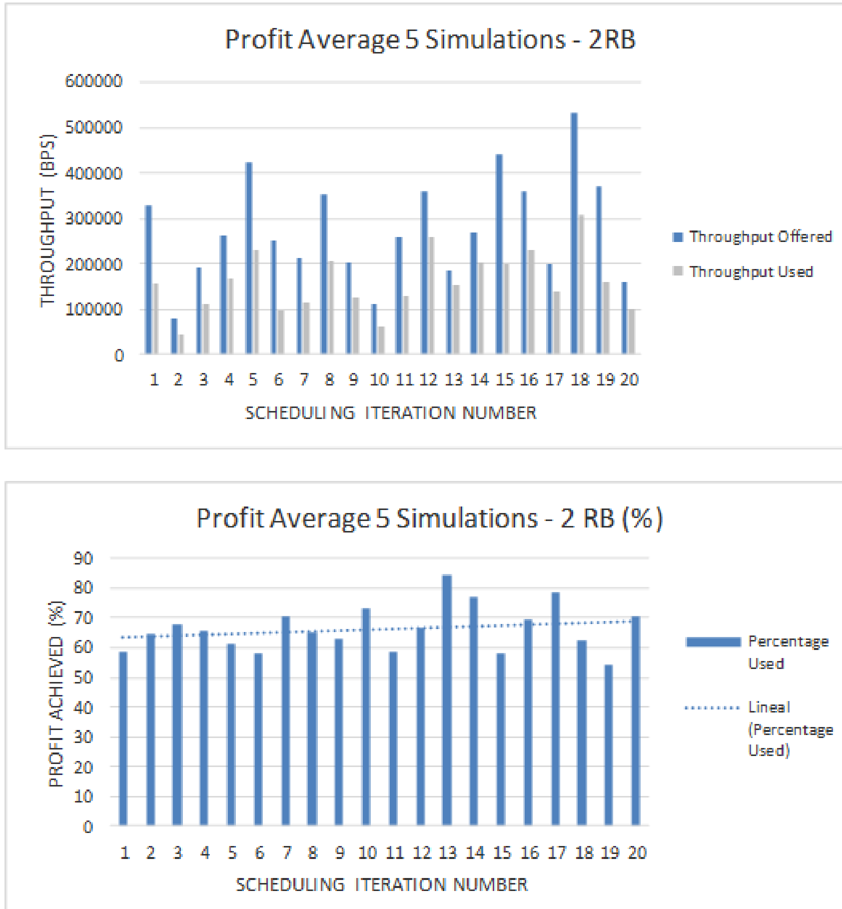


Figure 10: Profit Simulation First Scenario 2RB.

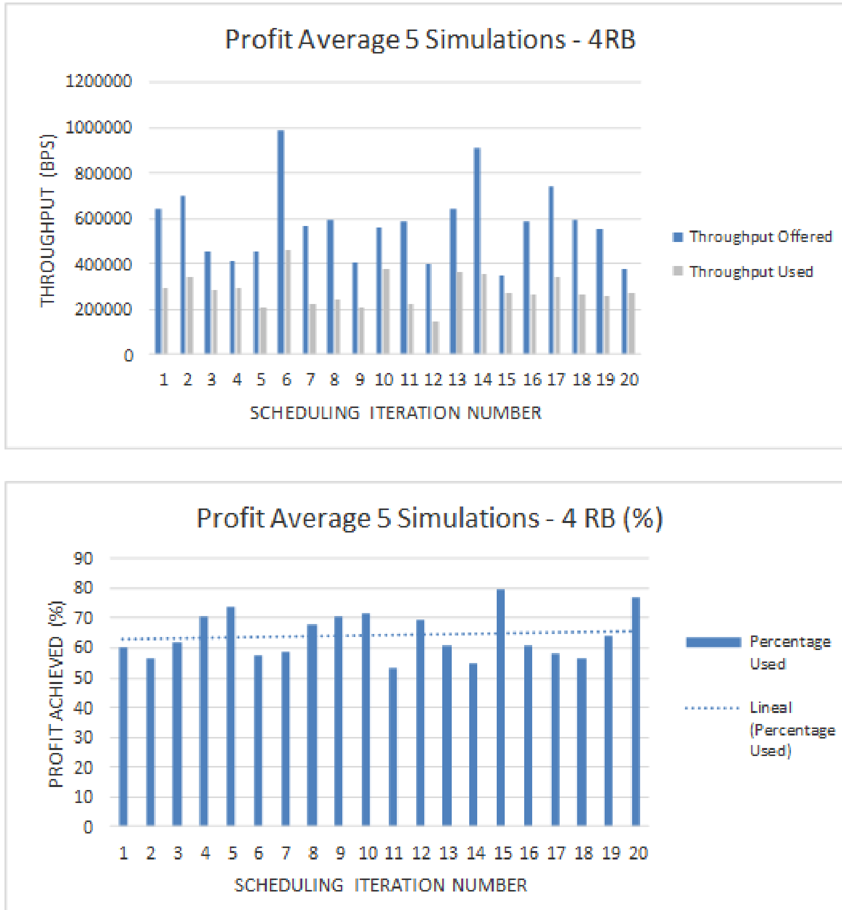


Figure 11: Profit Simulation First Scenario 4RB.

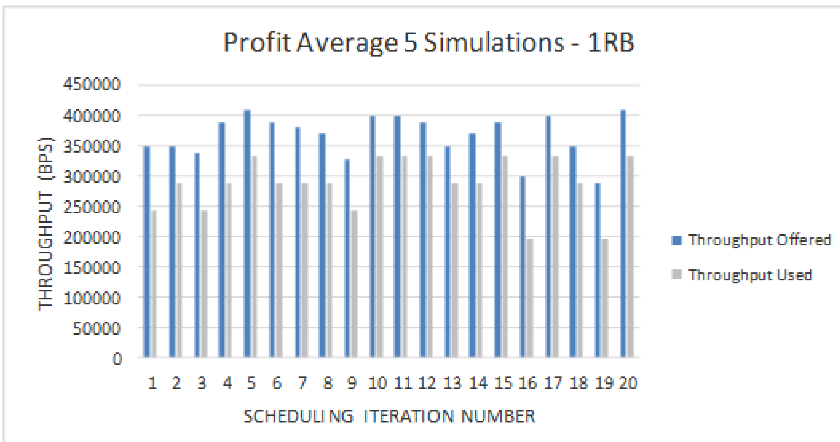
The idea of the second scenario was to verify that the efficiency of the allocation only depends on the relation between the channel conditions and the specifications of the flows at each moment. To do it, a second scenario with more strict traffic in terms of delay (HDTV, Real Time Gaming and Video Calls) was created and the channel conditions were modified giving users high offered rates by just reporting 64QAM-CQIs, from 10 to 15, with equal probabilities each of them. No power nor fairness controls were implemented and the buffer queue size was still set at zero. In this second scenario also the three different cases of having 1,2 and 4 RBs were considered.

The results obtained with the different combinations of RBs available are summarized

Flow id	User id	VN id	Service	Arrival Rate	Delay Budget
1	1	1	HDTV	30000 kbps	280 ms
2	1	2	RT Gaming	10000 kbps	30 ms
3	1	3	VideoCall	1000 kbps	130 ms
4	2	1	HDTV	30000 kbps	280 ms
5	2	1	RT Gaming	10000 kbps	30 ms
6	2	2	VideoCall	1000 kbps	130 ms
7	3	1	RT Gaming	10000 kbps	30 ms
8	3	3	VoIP	80 kbps	80 ms
9	3	3	HDTV	30000 kbps	280 ms

Table 4: Second Scenario Configuration.

in the graphs below.



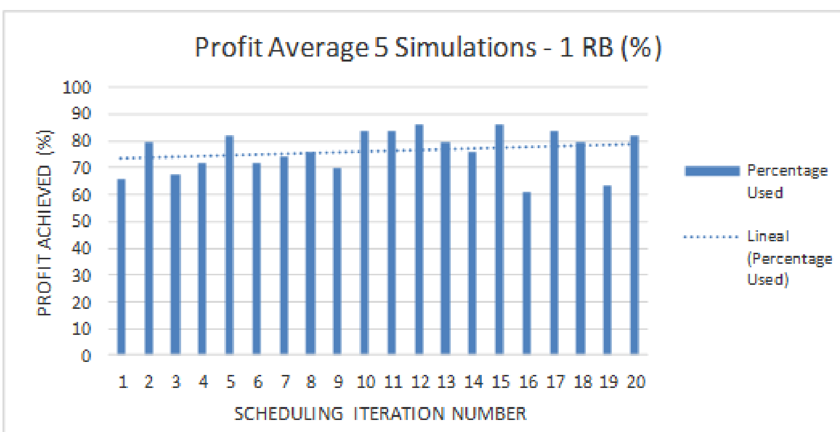
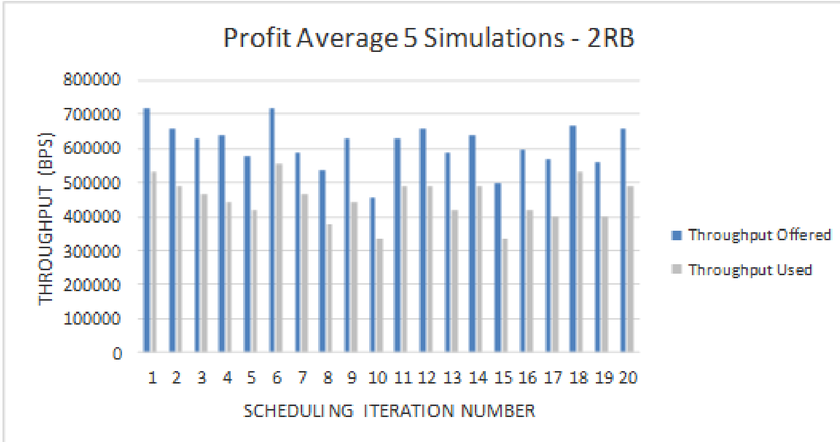


Figure 12: Profit Simulation Second Scenario 1RB.



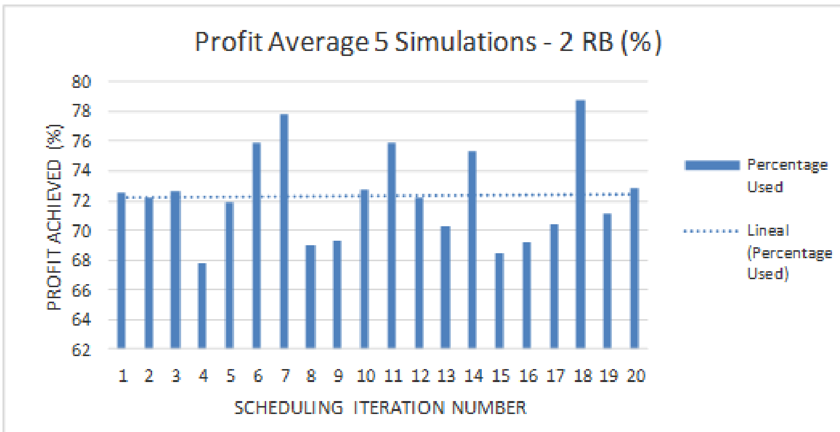


Figure 13: Profit Simulation Second Scenario 2RB.

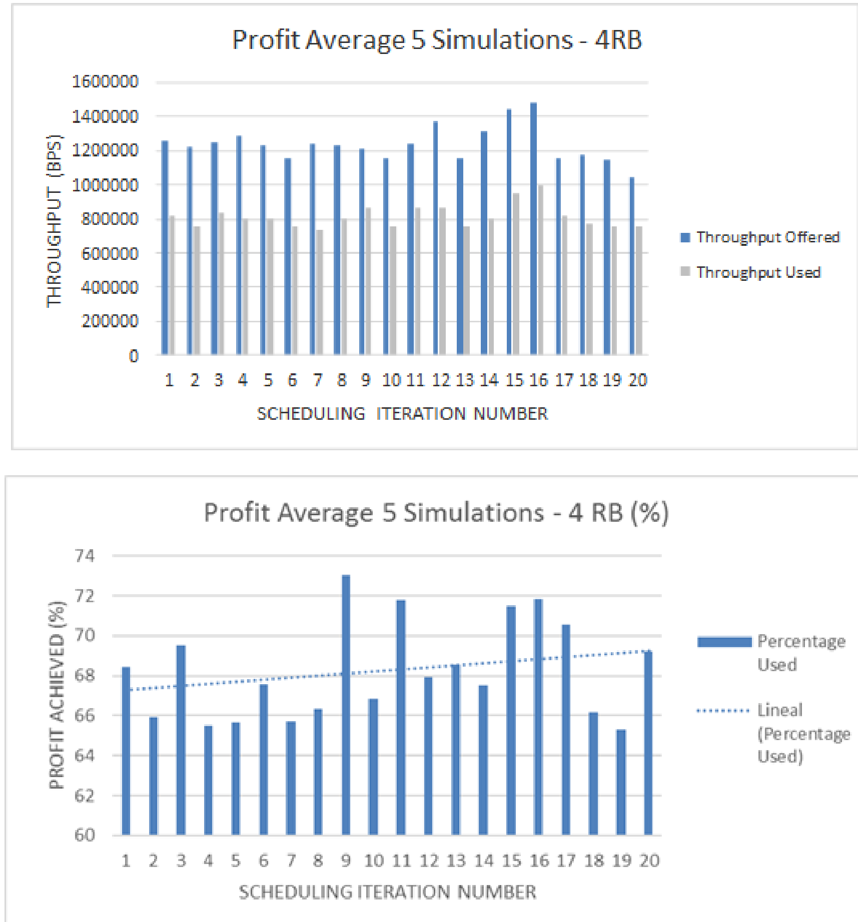


Figure 14: Profit Simulation Second Scenario 4RB.

As can be observed, the profit obtained by the flows only depends on their requirements and the state of the channel at every moment. When the users are perceiving better channel conditions and the throughput they need to transmit the data is big, high levels of efficiency can be achieved. However, in normal scenarios like the first one with not much throughput needed, the average profit that the selected flows obtain for the channel is not very high.

Another conclusion is that the profit does not depend on the number of RBs assigned to the flows, whenever the data needed to transmit is enough to fill all of them. Every RB has completely independent channel conditions, and thus the probability of having more or less profit, as the BW coherence selected is 12 sub-carriers, is just a matter of each individual RB.

In this case the channel conditions have been forced (in the second scenario), to find a situation in which the throughput requirements match the offer. Despite the fact that in real situations the channel will be completely random, the fact that in the future base stations will have a very big number of virtual networks with a very broad list of diverse requirements, will make the probabilities of finding flows matching channel offers higher, making this way possible the achievement of high levels of efficiency. Just because, even though the probabilities of obtaining good profit for each flow will be the same (considering absolutely random channel condition), the fact of improving the number of experiments, augment the probability of achieving good results.

To increase even more the efficiency aiming to find good levels, a third scenario was created where the services of the first scenario are considered but the required delay between the BS to the UE was reduced significantly. In the previous experiments we had considered a gateway to base station (P-GW to eNB) delay of 20ms as suggested in the standard. However, as this delay is dependent on the operator implementation, a scenario with much more strict delay and thus, higher throughput required, was analyzed.

Flow id	**User id**	**VN id**	**Service**	**Arrival Rate**	**Delay Budget**
1	1	1	HDTV	30000 kbps	40 ms
2	1	2	File Transfer	6000 kbps	40 ms
3	1	3	VoIP	80 kbps	20 ms
4	2	1	HDTV	30000 kbps	40 ms
5	2	1	File Transfer	6000 kbps	25 ms
6	2	2	Web Browsing	5000 kbps	30 ms
7	3	1	RT Gaming	10000 kbps	15 ms
8	3	3	VoIP	80 kbps	20 ms
9	3	3	VideoCall	1000 kbps	25 ms

Table 5: Third Scenario Configuration.

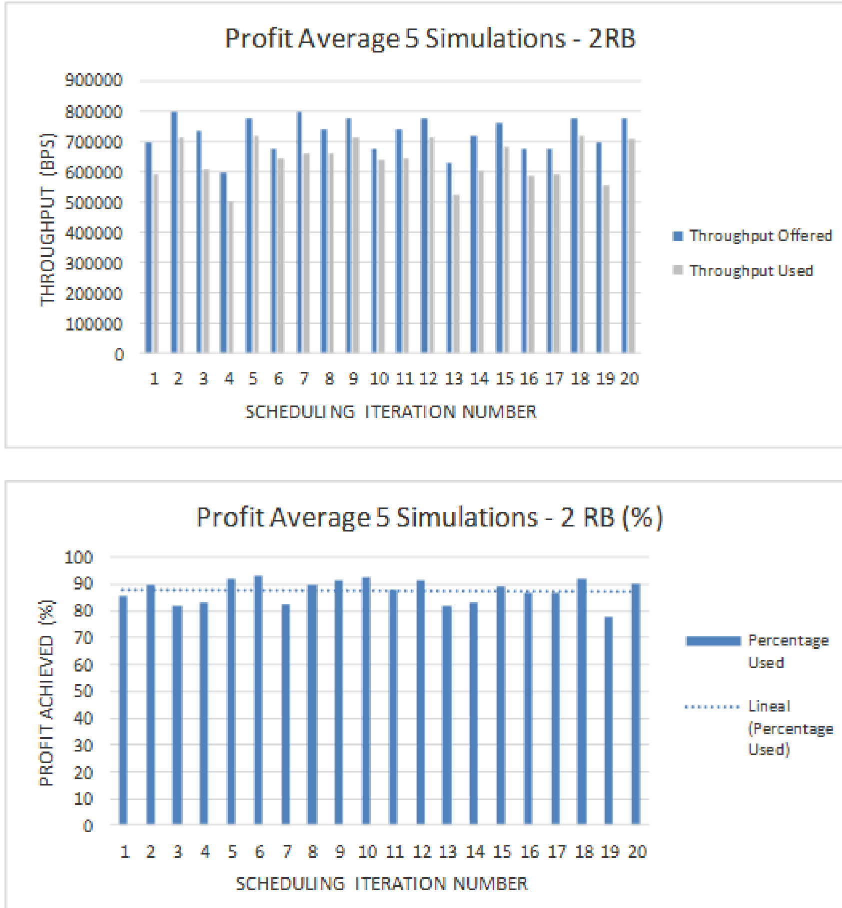


Figure 15: Profit Simulation Third Scenario 2RB.

Although having profits of 100% is almost impossible as achieving this rate would mean that in a specific moment, the selected flow just have the amount of data needed to transmit so that according to the delay required, the throughput resulting matches exactly the throughput that the flow is being offered at that moment, when having strict requirements and good channel conditions, profits around 90% can be easily achieved.

With these graphs the first set of experiments is concluded, having verified the good performance of our algorithm and the high levels of efficiency that can be obtained depending on the diversity of requirements in the virtual network environment and the channel conditions of the subscribed users at each moment.

2.3.2.2 Throughput Simulations

The next set of simulations were performed with the objective of comparing the performance of our solution with respect to other algorithms broadly used in literature. The first algorithm analyzed was Maximum SNR. Maximum SNR basically works by checking the user with best channel conditions at every moment and allowing him to transmit. One can clearly see even before performing any simulation that this technique has several problems. First of all and most important, it is not efficient, as having good channel conditions does not necessarily mean having a lot of data to transmit. This way the user with best channel state may have little data to transmit, making the decision of allocating all the resources to him very inefficient. Moreover, this user with best channel conditions can have several flows. Thus, a specific way to select which flow is the one having more priority to transmit among them has to be designed. This last issue happens with all the algorithms used in cellular communications until now, they are user based, not flow based, being this way not suitable for virtualization. In the same direction, Maximum SNR is almost the opposite of a fair algorithm, discarding the users with bad channel conditions regardless of their arrival rates, queues or requirements of their services and not providing isolation, which is the most crucial point for virtualization. If the fact that it does not ensure the satisfaction of the QoS requirements because it does not check if the channel offering is enough for the requirements of the selected user is added, we can conclude by saying that Maximum SNR is not an appropriate algorithm for virtualization and in particular, for solving our problem.

However, we created a modified version of the Maximum SNR algorithm to make it more flexible, where the SNR used is not by users but by combination of flows to compare its performance with our algorithm in terms of throughput achieved. From all the possible flow combinations (exhaustive research), the one having best channel offering is selected. As in the traditional max SNR, if the combination with best channel has low data to transmit, the efficiency will be very low and the resulting throughput will be low.

In an environment with relaxed conditions the bad results of Maximum SNR can not be perceived that well. For these reason, we have decided to perform the simulations in the more strict scenario (i.e. third scenario configuration), for the case of having 2RBs available to distribute, but allowing the users to obtain CQI reports from all the possible range (1-15). From the different simulations two results are analyzed. The first one is the throughput achieved, meaning the amount of data transmitted every ms as a result of the scheduling decision, and the second one is the useful throughput achieved, meaning the amount of data satisfying QoS transmitted every ms as a result of the scheduling decision. The results are summarized in the graphs below. The

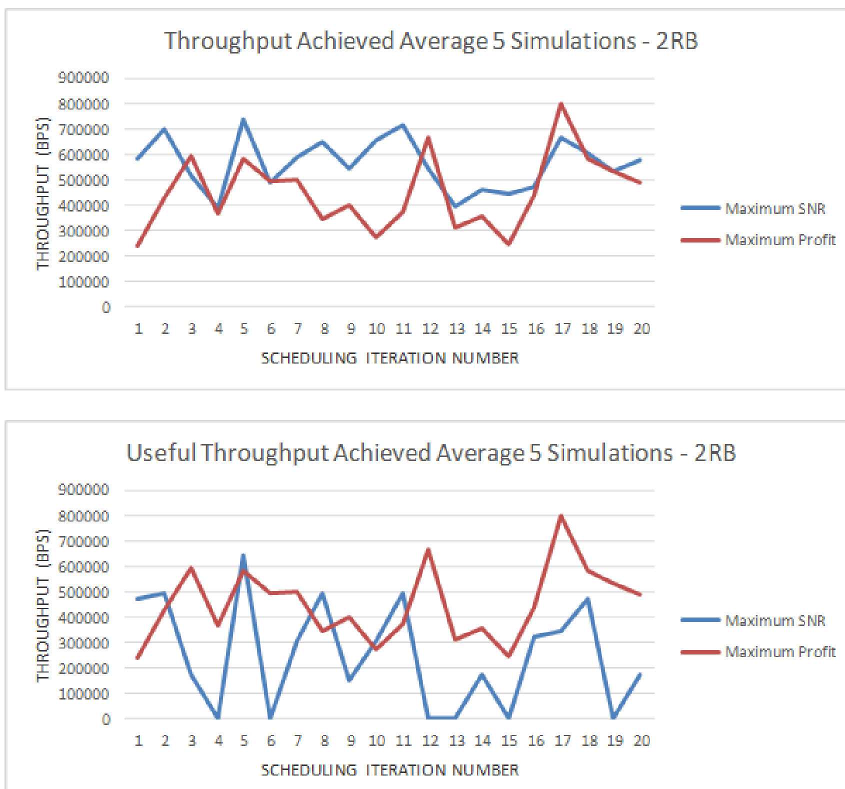


Figure 16: Comparison Performance Proposed Solution.

second algorithm analyzed was Weighted Fair Queuing (WFQ), which, differently from Maximum SNR, instead of considering the instantaneous channel conditions from each user to perform the scheduling, distributes the available resources proportionally between all users according to a specific weight assigned to each of them. In traditional WFQ, where the resource to distribute is time, the weight assigned to each user usually is the product of the queue size and the required rate. This way every user receives a proportion according to the assigned weight and the total weight computed as a sum of all the users weights.

$$RateAchieved_f = LinkRate \cdot \frac{w_f}{(w_1 + w_1 + ... + w_F)} \tag{29}$$

Usually WFQ is implemented in wired networks where the link rate is constant and equal for all the flows. In cellular communications, the channel rate is different for all the users and the partition of the link can not be done in all-size parts. In our case, as we already had designed a measure of each flow requirements that was *Throughput_required*, we created a Weighted Fair Queuing in which the weights were the following:

$$Weight_f = \frac{Throughput\_required_f}{Throughput\_required_1 + ... + Throughput\_required_F} \tag{30}$$

We realized that the results of this weights are very small fractions that can not be translated into RBs as we cannot assign half RBs, quarters of RB, or smaller parts of one RB. This way we created an adaption of the WFQ for the RB case in which the total weight is decreased progressively until all the RBs are filled with different flows according to their weight priority.

The scenario used to compare WFQ with our design scheduler is the same as for the Maximum SNR case, and the measures taken are also the same, throughput achieved and useful throughput achieved. The results are summarized in the graphs below.

The results for both scheduling algorithms (WFQ and MSNR) drive us to the same conclusion. In the MSNR case, the fact of giving the RBs to the flows with best channel conditions results in very high throughputs achieved, as for this scenario case

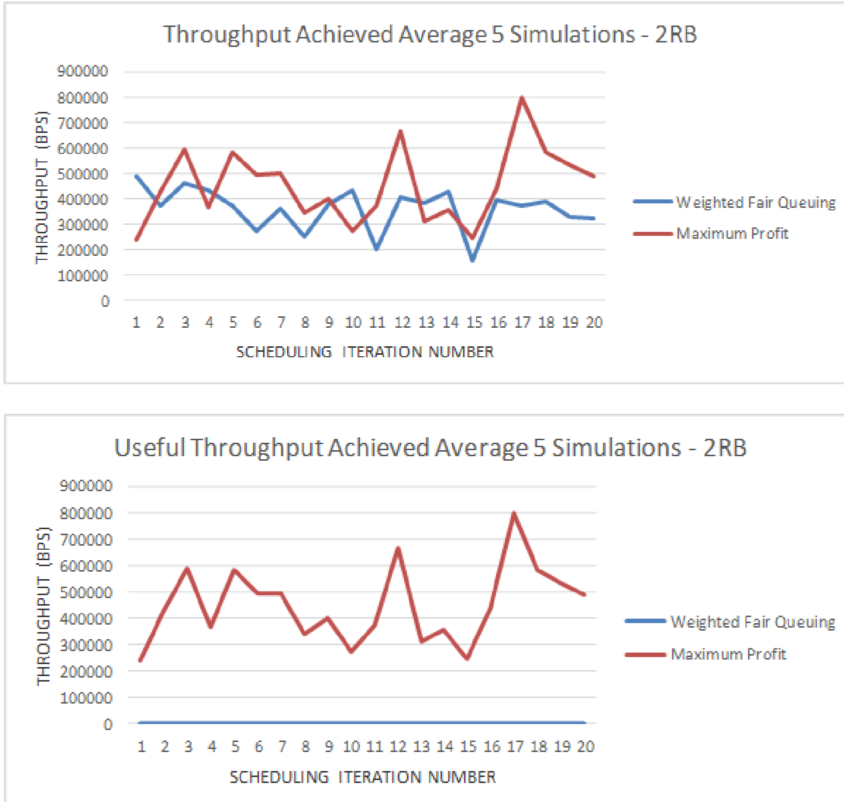


Figure 17: Comparison Performance Proposed Solution.

all the flows had sufficient data to transmit. However, what can be seen in the useful throughput achieved graphs is that in general, there is a lot of data transmitted in every scheduling decision, but the big majority of this data is not satisfying its QoS delay requirements, which means that these data, knowing that the services deployed are real time, will be discarded in reception, making the algorithm inefficient not only in terms of throughput but also in terms of power, using it for transmissions that will be discarded for not satisfying the requirements. For the WFQ the results are the same, although good throughput is achieved (not as much as with SNR), absolutely none of it is useful, due to the fact that, giving just a fraction of what every flow needs to make it more fair, makes that this fraction is not enough for the satisfaction of the requirements, making the transmission useless.

On the other hand, our algorithm not only achieves higher throughput than the other mechanisms but absolutely all this throughput is useful, making the transmission incredibly efficient. When considering fairness the throughput will not be that high but more isolation will be provided. The MNO has the flexibility to decide how to manage the trade-off.

2.3.2.3 Fairness Simulations

In this set of simulations, the objective was to verify the performance of the fairness-control mechanisms developed for our solution. The first scenario presented is considered but now not paying attention to the profit but doing it to the fairness achieved. One of the goals of our solution was providing isolation between different entities, virtual networks, flows and users. So in this section, scenarios in which fairness-control is not implemented and scenarios with fairness-control techniques will be compared. Figure 18 and Figure 19 show the evolution of the fairness index at a virtual network level. In the first one, no fairness control has been implemented. It can be seen how, while virtual network 1 is achieving high profit levels, virtual network 2 and more specially virtual network 3 remain with very low values of profit achieved. This is due to the fact that, while virtual network 1 provides services of HDTV and Real Time Gaming requiring high levels of throughput, which may match the high levels of channel quality, providing high profits, VN3 is offering voice and video services with much lower data rates and not that strict requirements, which may not be filling the channel offer for the current conditions. Observing that VN3 has low levels of profit, one option is to apply a fairness control just to VN3, allowing it to transmit although having low profit levels. But doing these kind of adjustments manually is not an ideal solution, as an analysis of the specific solution has to be performed in advance to determine how much isolation is needed. In the scenario of Figure 18, according to the average channel conditions and the requirements of virtual network 3, a bound of 5% was appropriate.

To manage the fairness in an adaptive and dynamic way, working with the standard deviation parameter is proposed. The standard deviation quantifies the amount of variation or dispersion of a set of data values. A low standard deviation indicates that the data points tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the data points are spread out over a wider range of values. Computing the standard deviation of the fairness index of the different entities (virtual networks, users or flows) and taking scheduling decisions restricting its value to be lower than a bound is an appropriate solution for achieving efficient and dynamic fairness-control.

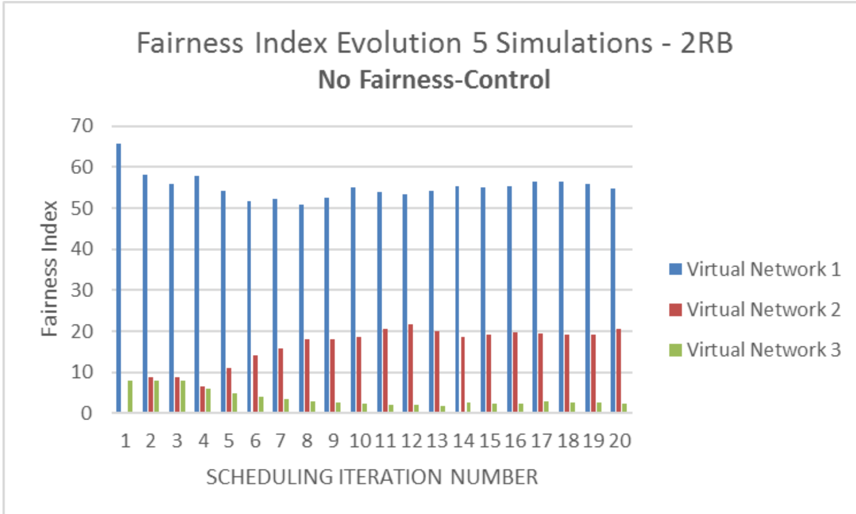
$$\sigma = \sqrt{\sum_{i=1}^{n}(Fairness_i - Fai\bar{r}ness)^2}. \tag{31}$$

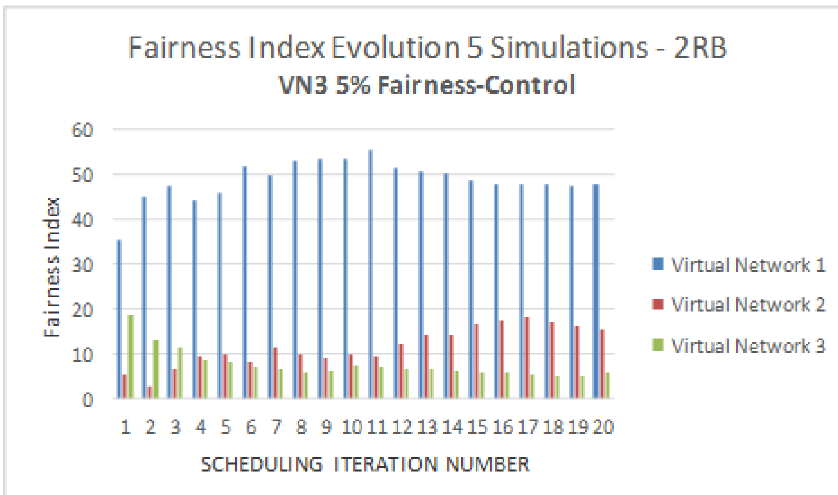
Finally to analyze fairness at user level, as the users profit were very homogeneous, we modified the channel conditions giving all of them good channel quality, with CQI reports from 10 to 15. This way, as can be seen in the second graph, the third user is the one taking more profit from the channel and not allowing the other two users to transmit. In this situation, a fairness-control mechanism applying standard deviation bound was applied and good isolation levels between users is achieved.

2.3.3 Conclusions

As the obtained results for the simple scenario were really good and high levels of throughput and efficiency were obtained, the next step of the project was to broaden it to more complex scenarios, in particular considering the multi-RAT capability of the wireless hypervisor. One of the main characteristics of 5G cellular systems is the diverse environment of different network technologies coexisting and giving service to the same users. So our wireless hypervisor needed to provide optimum cross-technology performance. The main issue, which will occupy our attention during this chapter, is the coordination of the two principal radio access technologies: 802.11 and

LTE. As LTE was already deeply studied in the chapter before, the objective of the following point is the description of IEEE 802.11, and the capabilities that it offers to the network virtualization requirements.





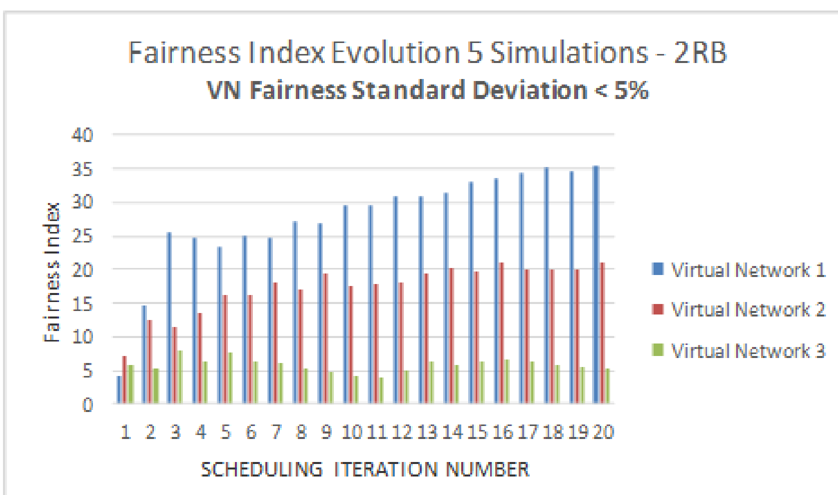


Figure 18: Fairness Performance Virtual Network Level.

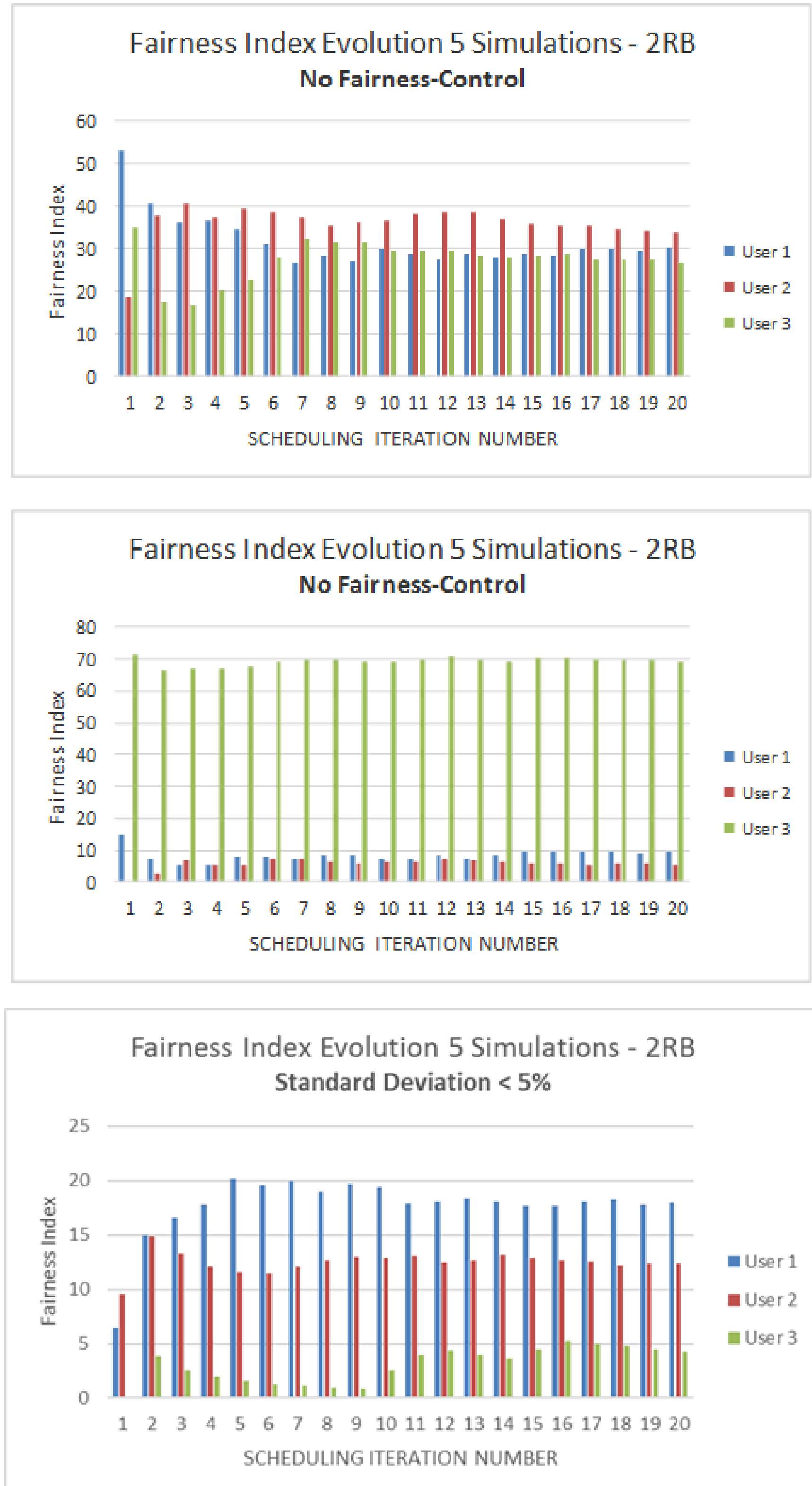


Figure 19: Fairness Performance User Level.

CHAPTER III

MULTI RADIO ACCESS TECHNIQUES

It is a clear fact that networks are becoming every day more heterogeneous as we move towards next generation 5G. The integration between all the different radio access techniques (RATs) sharing the air resources and creating this diverse wireless environment will be a key feature in the design of future communication systems. 5G-enabled devices will not only support a potentially new 5G standard, but also numerous releases of 4G LTE, several types of the wireless local area network (WLAN) IEEE 802.11 family of standards, commercially certified as WiFi, wireless metropolitan area network (WMAN) technologies like WiMAX and perhaps direct device-to-device (D2D) communication, all across a broad spectrum band.

It is important to notice that not all these technologies can equally benefit from wireless virtualization. The benefits of virtualization are more apparent when the supported data rate and the supported number of users are relatively high, leaving room for possible sharing of resources on the infrastructure [16]. Thus, 802.11 WLAN technologies and cellular technologies, which satisfy the requirements of high data rate and high number of users, are the main focus of most active research in wireless virtualization. Having analyzed the key design issues for cellular technologies, and having implemented a solution for the LTE-based scenario, the objective of this chapter is to extend the wireless virtualization studies for the IEEE 802.11 case and broaden our solution for hybrid networks where both 802.11 and LTE are shared by the service providers and their respective virtual networks.

3.1 IEEE 802.11 Challenges

Differently from cellular technologies, which have a natural affinity for virtualization due to their advanced quality-of-service (QoS) support, multiuser multi-access scheduling and network management capabilities, 802.11 WiFi technologies have a less elaborate control and management framework due to the plug-and-play nature of its setup. IEEE 802.11 Wireless LAN (WLAN) standard was developed with the objectives of low cost, simplicity, robustness against failures, and easy deployment. This philosophy, thanks to which has witnessed a phenomenal growth, being nowadays widely deployed and used around the globe, makes it contain a certain group of properties that suppose a challenge when trying to apply virtualization.

The first and most remarkable one is that IEEE 802.11 is based on CSMA/CA, a distributed medium access mechanism where the different users or stations fight between them to achieve the channel, listening continuously the medium to prevent collisions. This distributed technique, in which there is no scheduling, and are the stations the ones that directly compete for the resources, seems to be going against the efficiency-based centralized philosophy of Software Defined Networking. Moreover, and also due to the initial desire of easy deployment and robust performance, the native IEEE 802.11 does not support QoS differentiation. All the stations fight for the channel in an equal-possibility scenario, where all of them have to wait the same inter frame time, and a random back-off time from a same-length possibilities interval. Achieving granular QoS satisfaction, however, is the essence of virtualization, which aims to differentiate virtual network flows requirements. All these reasons, and some other problems such as not supporting channel quality information feedback, going against our idea of maximizing the channel profit, make our desire to adapt our first solution to the new concept of radio access technology more challenging.

- Native IEEE 802.11 is based on the medium access mechanism CSMA/CA that is not centralized, going against the philosophy of SDN.

- Native IEEE 802.11 does not support granular QoS which is the essence of virtualization, which aims to differentiate virtual network flows requirements.
- Native IEEE 802.11 does not contain channel quality information feedback, going against our idea of maximizing the channel profit.

According to the best of our knowledge, this is the first proposal considering wireless virtualization in an heterogeneous environment at a flow level and we believe that the chance of being able to do it with the capabilities that an SDN architecture offers will allow us to achieve the efficiency and flexibility needed and that this new idea will open a new branch of research for the next generation cellular networks. To do it, all the challenges introduced in this first section of the chapter will need to be resolved.

IEEE 802.11	**3GPP LTE**
Designed to work in free spectrum: Not highly optimized: Packet headers and contention time give overhead. - Access is by CSMA/CA. - AP has no “privilege” over stations, except PCF or HCCA. - PHY is packet oriented –sync on each packet. - PHY provides a single channel with a single modulation for each packet. - PHY interface to the MAC is by a Protocol Data Unit (PDU).	Protocol designed to maximize utilization of expensive spectrum. Very complex and highly optimized. - All client access is scheduled. - Base Station has absolute control over network device operation. - PHY operates continuously –sync is interspersed. - PHY provides multiple channels simultaneously with varying modulation. - PHY interfaces to the MAC with a Transport Block.

Table 6: Comparison IEEE 802.11 and 3GPP LTE

3.2 HCCA: Centralizing CSMA/CA

To enhance the support of QoS, the IEEE 802.11e group developed a new protocol that proposes differentiation mechanisms at the medium access control (MAC) layer. It uses a new medium access method called the hybrid coordination function (HCF) that combines a contention-based enhanced distributed coordination function (DCF) access mechanism (EDCA) and a controlled HCF channel access mechanism (HCCA) in a single function providing much more flexibility in the scheduling of resources and increasing quality of service.

EDCA is based on CSMA/CA, with the peculiarity that stations can choose between four different Access Categories, according to the type of traffic that they are providing, receiving this way more or less priority according to their requirement, by having modified contention parameters such as TXOP interval size and CWs. This first technique is not suitable for the aim of our project as it does not satisfy the software defined network centralized philosophy. We could try to compute the achievable throughput of the different stations with the more than well-known Bianchi equations [4], based on the collisions probabilities, but we would not be able to guarantee a specific data rate, making virtualization impossible to implement.

However, the second technique proposed (HCCA), offers us the needed flexibility that we are looking for. In HCCA a station called Hybrid Coordinator (HC) centralizes the medium access process. One can think of it as a variant of the TDMA. In HCCA, the time is considered as continuous super-frames with a contention-based phase (CP) and a free-of-contention phase (CFP). The Hybrid Coordinator has highest priority during both CFP and CP. In CFP the stations can only transmit when they receive a polling request from the HC. Whenever the HC wants to transmit, it just has to do it. In CP, however, all the stations are trying to transmit. The priority, thus, is given by the fact that the HC only has to wait for a period of time PIFS instead of DIFS, as the rest of stations (being PIFS<DIFS), to transmit. Once it transmits, the HC

may allocate TXOPs to itself to initiate MSDU Deliveries whenever it requires, or sending a QoS-Poll to any of the other stations, giving it the permission to transmit during the specified TXOP [17].

HCCA is the most centralized medium access technique that 802.11 technologies are offering, being this way the most suitable for our project as this centralization improves radically the efficiency of the system and matches perfectly with the SDN philosophy on which our project is based. Although 802.11 does not natively lie on that HCCA capability and a lot of technology in the market still does not support it, we believe that on the new generation cellular technologies the trend will move towards centralized strategies and 802.11 will focus more on this kind of techniques. Despite native 802.11 was operating in a completely distributed way, the complexity of the new arising traffic as well as the demands for more specific QoS and the trend towards the well-known infrastructure as a service will make the next generation networks to look for centralized architectures improving efficiency and providing the needed granularity at all levels.

3.3 IEEE 802.11 Scenario: Problem Formulation

Suppose the scenario in which a base station (or access point by now) has different virtual networks allocated, providing services to different subscribed users around the cell. Suppose that both users and access point support IEEE 802.11ac[7][13] in terms of physical and MAC layer and IEEE 802.11e traffic priorization mechanisms, including HCCA. Considering only the Contention Free phase of the super-frames, the base station is the one that centralizes the time division between flows to deliver their data to the users. As the other stations (or users) are not supposed to transmit, the HC does not need to wait for a PIFS time in every action and the efficiency can be maximized. For the physical transmission 802.11ac is the one playing an important role, allowing the achievement of high throughputs and providing mechanisms to improve efficiency scheduling such as channel information reporting. Differently from the LTE case, the channel bandwidth will be fixed, giving all the data-useful sub-carriers to the selected flow at each moment, being this way the time the only variable to distribute. As only downlink is being considered, the station will need to determine how much time will dedicate to the transmission of each flow. This way, the time scheduling will be performed by allocating time slots called transmission opportunities (TXOPs) to different flows of the network. The TXOP time required by each flow will vary depending on the channel conditions at each moment, the data available to transmit and the requirements to be satisfied.

The scheduling decisions are taken every intervals of time called service intervals, SI, which are equivalent to the transmission time intervals of LTE.

The objective, thus, is to give access to the different flows of the base station, with a combination of TXOPs that maximize the profit of the channel resources while satisfying their data rate requirements and providing fairness between all the parties.

3.4 IEEE 802.11 Scenario: Proposed Solution

Once the problem has been introduced, in this section a solution will be proposed, adapting the algorithm developed for the LTE case to IEEE 802.11.

Having defined every flow allocated to the base station with a specific arrival rate and a QoS delay requirement to be guaranteed, the data available to transmit at the next decision can be defined as:

$$Data\_required_f = Arrival\_rate_f \cdot SI \quad (+buffer) \tag{32}$$

And in the same way we had done for LTE, the throughput required can be computed.

$$Throughput\_required_f = \frac{Data\_required_f}{QoSDelay_{\min,f}} \tag{33}$$

Differently from LTE, IEEE 802.11 is packet based and synchronization has to be implemented, forcing us to consider all the overheads to make our solution standard compatible. Two packet sizes will be considered, the nominal size referring to the size occupied by only useful data (and symbolized with ' in the equations) and the real size containing all the extra information for signaling and overheads. Considering equal packets of maximum data size:

$$Packets_f = \left\lceil \frac{Data\_required_f}{MaxLength'} \right\rceil \tag{34}$$

To increase channel efficiency the last packed difference of size will be also considered.

$$LastLength' = Data\_required_f - (Packets_f - 1) \cdot MaxLength' \tag{35}$$

As 802.11ac inherits the channel quality reporting mechanism from 802.11n, which is used for high-throughput applications such as MIMO, it can be used in the same way as it was done for LTE to compute the throughput offered by the channel.

$$Throughput\_offered_{u[f]} = i_f \cdot \frac{1\ sym}{tt_{OFDMA}} \cdot BW_{WiFi} \cdot Code\ Efficiency_{u[f]} \cdot Modulation_{u[f]} \tag{36}$$

Where i_f is the selection index, being 1 if the flow is accepted and 0 if not, tt_{OFDMA} is the transmission time of one OFDMA symbol, BW_{WiFi} is the number of subcarriers assigned to the WiFi channel, and code and modulation are the result of the channel state reports mechanism that selects the optimum data rate for every specific user. The time required to transmit all the data information can be defined as [3]:

$$\begin{aligned} TXOP_f = (Packets_f - 1) \cdot (&\frac{MaxLength}{Throughput\_offered_f} + 2SIFS + ACK) \\ +(&\frac{LastLength}{Throughput\_offered_f} + 2SIFS + ACK) \end{aligned} \tag{37}$$

$$s.t. \qquad TXOP_f \leq TXOP_{max} \tag{38}$$

Finally, the profit computation, which indicates how appropriate is the allocation of the medium to the flow can be described in the same way it was done for LTE.

$$Profit_f\,(\%) = \begin{cases} 0\ \ if\ \ Throughput\_offered_f = 0 \\ \\ \frac{Throughput\_required_f}{Throughput\_offered_f} \cdot 100\ \ otherwise \end{cases} \tag{39}$$

To conclude, the optimization can be written as:

$$\max_{i_f} \sum_{f \in F} i_f \cdot Profit_f \tag{40}$$

$$s.t. \quad \sum_{f \in F} \frac{TXOP_f}{SI} \leq \frac{dot11CAPRate}{64[\mu s]} \tag{41}$$

Where the constrain reflects the bound imposed by the 802.11 standard of maximum contention free time. SI has to be sub-multiple of the 802.11 inter beacon time, and will be set with the same value as LTE TTI to synchronize both RAT decisions. Finally, power and fairness control can also be implemented if desired.

$$Total\_power \leq P_{WiFi\ RAT} \tag{42}$$

$$\varphi_1 \leq Fairness_{u,f,s} \leq \varphi_2 \quad \forall\, u, f,\ s \tag{43}$$

According to the best of our knowledge our proposal is the first one that introduces WiFi virtualization by exploiting the opportunities given by IEEE 802.11 HCCA. In our opinion, the centralized philosophy that HCCA is offering should be further investigated and we strongly believe that, as soon as virtualization interest for SDN architectures augments, the research community will focus more on all those centralized mechanisms that increase the network efficiency and its global performance.

3.5 MultiRAT Environment: Problem Formulation

Once both LTE and WiFi have been modeled individually, consider a scenario where both technologies are merged together, so that the different flows from the base station can be transmitted using one or the other radio access technology.

Assuming that only one technology can be selected by each flow at every decision, the final optimization can be written as:

$$\max_{i_{f,rat}} \sum_{f \in F} \sum_{rat \in RAT} i_{f,rat} \cdot Profit_f \tag{44}$$

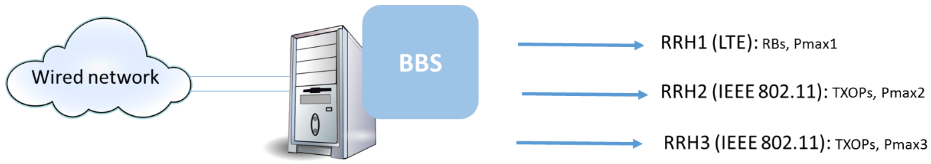


Figure 20: Multi-RAT scenario.

where:

$$Profit_f\,(\%) = \begin{cases} 0 \;\; if \;\; Throughput\_offered_f = 0 \\ \\ \frac{Throughput\_required_f}{Throughput\_offered_f} \cdot 100 \;\; otherwise \end{cases} \tag{45}$$

and:

$$Throughput\_offered_{f,rat} = \begin{cases} \frac{1\ sym}{tt_{OFDMA}} \cdot BW_{WiFi} \cdot MCS_{u[f]} \; if \;\; rat = WiFi \\ \\ \sum_{r=1}^{RB} i_{r,f} \cdot 7sym \cdot 12sc \cdot MCS_{u[f]} \;\; if \;\; rat = LTE \end{cases} \tag{46}$$

$$s.t. \quad \sum_{f \in F} i_{f,WiFi} \cdot \frac{TXOP_f}{SI} \leq \frac{dot11CAPRate}{64[\mu s]} \tag{47}$$

$$Profit_f \leq 100 \qquad \forall\, f \tag{48}$$

And the fairness and power control if desired:

$$\varphi_1 \leq Fairness_{u,f,s} \leq \varphi_2 \quad \forall\, u,\, f,\; s \tag{49}$$

$$Total\_power \leq P_{LTE\ RAT} \tag{50}$$

3.6 MultiRAT Simulations and Final Solution

To analyze the performance of the algorithm for the multi-RAT case, and the possible problems that could arise from the direct extension of the single-RAT scenario to the new technology adopted, some simulations were developed. A scenario with one BBS and two RRH was considered, where each RRH supported a different radio access technology, one supporting LTE and the other WiFi. For the WiFi one, as stated in the beginning of the chapter, 802.11ac amendment was considered. Both RRHs had the same coverage, allowing all users of the cell to connect to one or the other in-distinctively. Same as in previous scenarios, these users were subscribed to some of the services provided by the base station virtual networks and had different flows requiring to satisfy diverse QoS. To find the optimum allocation of flows to the different RATs, and to the resources available in these RATs, the optimization proposed in the previous section was solved by exhaustive research methodology.

In a first instance, all the possible assignments flow-RAT were computed, so that every flow could be allocated to LTE, to WiFi or not selected. For each combination in which LTE had been designed to some of the flows, all the possible distributions of LTE resources (i.e. RBs) among these flows were also calculated. All the distributions were analyzed to verify if the QoS requirements could be satisfied and compared to select the one giving the best result (i.e. maximizing the profit) at each moment.

For the flows allocated to LTE, the throughput offered to their users with the current channel conditions was computed from the last CQI report information received and compared to the throughput required by the flow. For the WiFi allocated flows, not only the throughput offered was computed with the latest MCS information, but the time needed for the information transmission was also calculated to verify if it could be fitted in the decision interval slot. To do it, all the data received by each flow was divided in MSDU packets as specified in 802.11ac, which were then attached in the data slot of the MPDU frame with their respective overheads. Knowing the throughput offered, the time to transmit the physical frame in which the MPDU was allocated was computed. With it, and the SIFS and ACK times the TXOP size needed could be finally obtained. If all the restrictions were satisfied, the profit achieved was compared to select finally the best combination for the moment decision.

A key point for succeeding with our algorithm was to achieve synchronization between RATs. In the second chapter it was reviewed that LTE performed decisions

IEEE 802.11 PARAMETERS	**VALUE**
Channel Bandwidth	80 MHz
Total Subcarriers	242
Data Subcarriers	234
OFDMA Symbol	3.6 us
Service Interval (SI)	1 ms
dot11CAPRate	64 us
MCS Feedback Periodicity	1 ms
MSDU Max Size	2304 B
MSDU Overhead	14 B
MPDU Overhead	40 B
PPDU Preamble	40 us
SIFS	16 us
ACK	14 B

Table 7: IEEE 802.11 Used Specifications.

on intervals of 1ms. The IEEE 802.11 HCCA philosophy however was a bit different, with a suggested scheduling [6] controlling allocations on-the-go dividing the time in SI. To synchronize the scheduling of both technologies, and to be able to compare them in equal conditions, the IEEE SI was fixed at 1ms, making it act paired to the LTE TTI.

The dot11CAPRate was set to 64us to test the performance limits of the scheduling algorithms and observing the HCCA scheduling channel utilization efficiency. All the defined range of modulations from the standard were considered and a bandwidth of 80MHz was chosen to obtain the enough flexibility needed to work under a 1ms decision rate. The rest of used specifications for the IEEE 802.11 added technology are summarized in Table 7.

From these first set of simulations some interesting results were obtained. On the one hand, 802.11ac, being characterized for using dense modulation schemes, arriving to 256-QAM, and broad bandwidths until 160MHz, provided the achievement of very high throughputs, allowing the transmission of long extensions of data in very small TXOPs, resulting ideal for working at TTI level, but also improving the performance of the global system.

On the other hand, however, these high throughputs achieved, were making the profit indexes of the flows decrease significantly, not allowing us to exploit at the maximum the resource efficiency that the heterogeneous environment was offering, and allocating all the flows to LTE technologies in every decision.

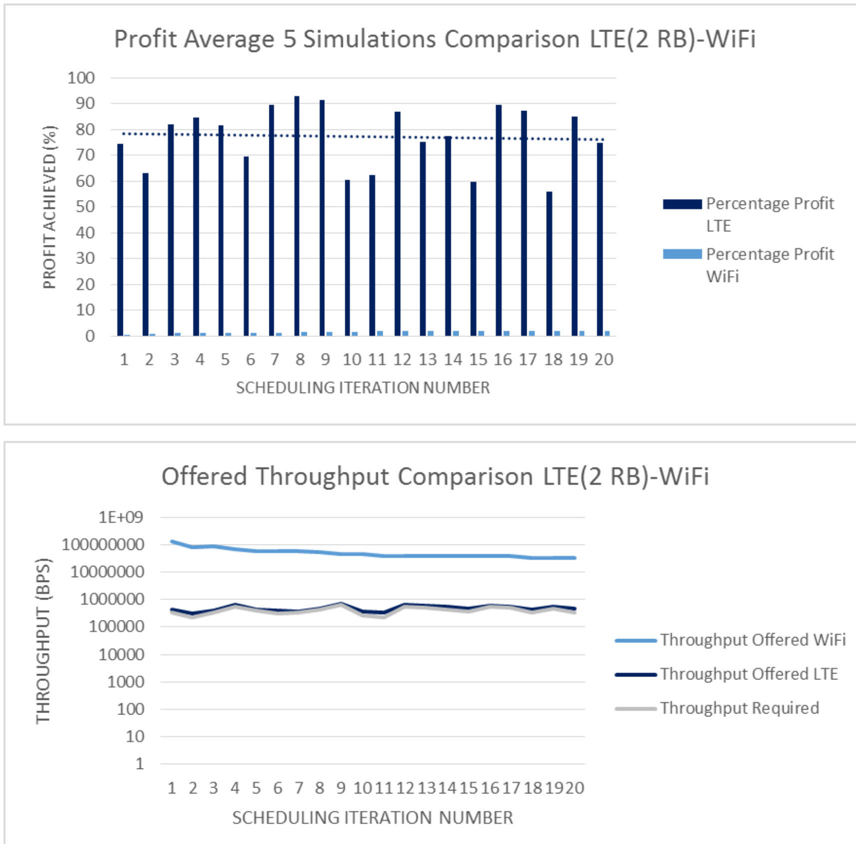


Figure 21: Profit and Throughput Comparison LTE-WiFi.

As the small profits obtained were showing that the requirements of the WiFi allocated flows were always being satisfied by far, a final solution in which the objective for WiFi was not to maximize the profit but maximize the time efficiency was designed. It was based on a two-steps optimization in which, in first instance, and giving priority to WiFi, the combination of flows best fitting in the available decision slot was selected and, in second instance, the best distribution of RBs among the rest of the flows was computed. A flow chart detailing the steps followed by this last proposed algorithm

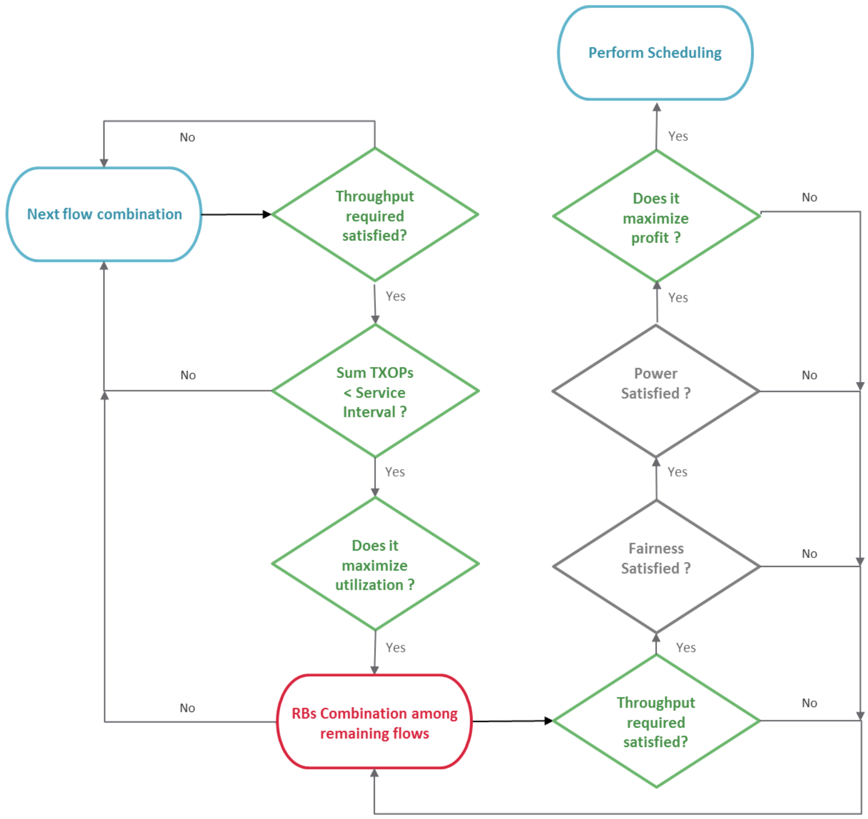


Figure 22: Two-Step Optimization Flow Chart.

can be found in Figure 22. The idea was the same as for the maximum profit solution proposed in the previous section, with the only difference that for the WiFi case, the best combination was not found in the one maximizing the flows profit but in the one maximizing the available slot-time (SI) efficiency. And instead of comparing both technologies at the same time, WiFi was given more priority, as it offered the best service to the flows, allowing them to transmit all the queue data in a single intent at incredibly high data rates. This last solution developed, could finally take advantage of the resource offering of both technologies and the global efficiency was finally maximized. The simulations performed with this last solution proved the optimality of the results in terms of efficiency at all levels, flexibility and fairness granularity. In terms of fairness, the updating and predicting algorithms remained the same as for the single-RAT case with the only peculiarity that the profit given

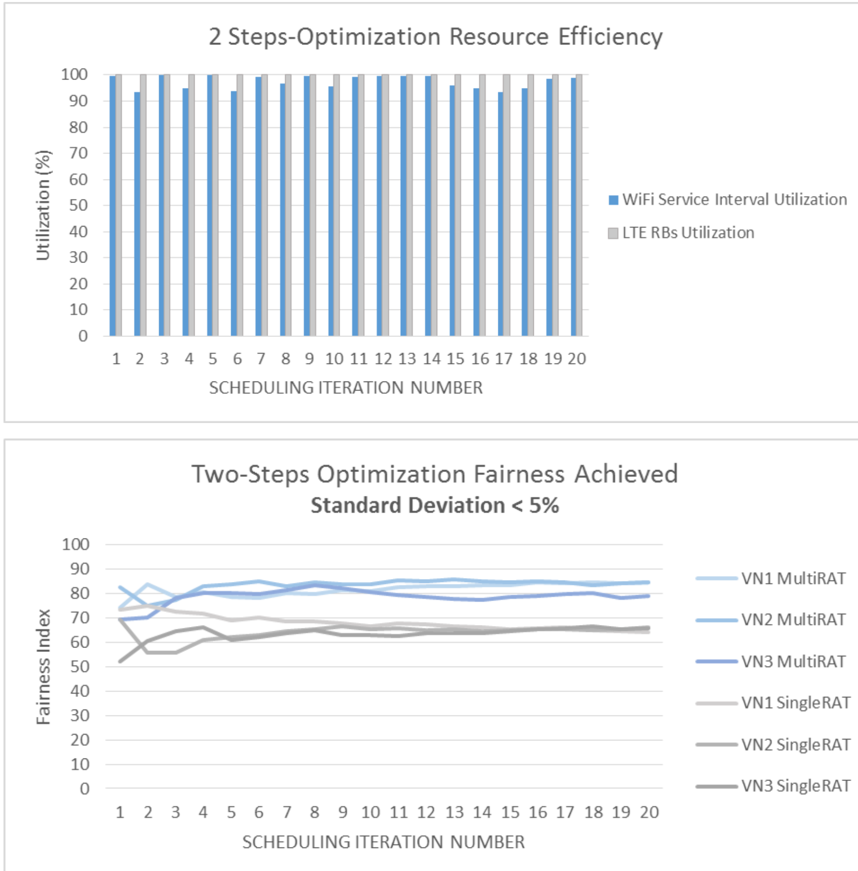


Figure 23: Efficiency and Fairness Final Solution.

to WiFi selected flows was defined as 100% due to all the advantages offered by this technology, already commented. And with these results, we concluded the thesis satisfied of having been able to design throughput-efficient and utilization-optimal algorithm allowing wireless virtualization for the heterogeneous networks of the next generation 5G cellular systems.

CHAPTER IV

CONCLUSION

The objective of this thesis was to move towards the design of a throughput-efficient wireless hypervisor, determining the optimal distribution of non-conflicting network resource blocks among virtual network operators, in terms of channel resources, power levels and radio access technologies, so that the data rate requirements demanded could be guaranteed and the global throughput efficiency maximized.

4.1 Contributions

Throughout the first part of the project, the motivation for the use of wireless network virtualization in the next generation cellular technologies was introduced and the general requirements and challenges for its implementation were reviewed. After analyzing the state of the art literature proposals, the key design aspects for providing a flexible but efficient solution were proposed. The implementation of our solution was started with the simple case scenario of one only base station, in which the optimal allocation of resources among flows was computed. The mathematical framework developed provided a complete analysis of the different parties requirements and the channel conditions at every moment, enabling coexistence, flexibility and isolation at all levels. All the theoretical results were accompanied by simulations deployed on a self-developed highly manageable and configurable object oriented based simulator, allowing to verify the performance in a highly diverse range of scenarios and environments. These simulations proved the optimality of the designed algorithm results and its clear improvement with respect to the currently used technologies.

In the second part of the project, the work was broadened to a multi radio access technology capable scenario in which IEEE 802.11 was considered, adopting scalability and providing heterogeneity, increasing the overall performance for the highly diverse and rapidly changing traffic patterns our solution was aimed to face.

By exploiting the opportunities given by the new standards developed as well as the capabilities offered by a SDN based architecture, a two-steps optimization was finally proposed. With the designed solution, not only the rates of the different virtual networks are satisfied, but isolation at all levels is guaranteed and the global performance in terms of throughput is maximized. For all these reasons we can conclude by saying that with this solution, the objective of the design and implementation of throughput-efficient algorithms towards wireless virtualization for 5G cellular systems has been finally achieved.

According to the best of our knowledge, this is the first proposal introducing throughput-efficient and utilization-optimal wireless virtualization for an heterogeneous environment with a flow-level granularity and we strongly believe that it will open a new path in research, bringing the community to investigate towards highly centralized mechanisms that can face the challenges imposed by the next generations of cellular technologies.

4.2 Next Steps

Although having set the basis for the design and implementation of wireless virtualization for 5G cellular systems, this is just the beginning of a very long journey and there are still a lot of challenges and opportunities remaining to be addressed.

In first instance and concerning directly the work developed in this thesis, it is interesting to mention the following next steps.

During all the thesis, we have defined a solution in which the only consideration and objective has been to achieve the optimal results. However, the computation of these optimal results sometimes can be complex. Once the optimal performance has been found, our next objective is to achieve algorithm optimality, i.e. finding way to obtain close-to-the-optimal results in a cost-effective manner. In this sense two factors play an important role, which are complexity and convergence.

To obtain a simplification of the proposed solution, we are currently considering two techniques, clustering and pattern recognition, aiming to solve the problem in linear time by relaxing a little bit the optimality of the results. Clustering basically means grouping different parties together to compute a single factor for all of them. In our case, the performance resulting of clustering flows, virtual networks and users with similar requirements is being investigated. Pattern recognition aims at finding repeated behaviors in the algorithm that can allow a reduction of the solution computation. Identifying that some flows with specific characteristics always are allocated in the same way or finding periodicities in the scheduling decisions can help us to reduce the number of operations performed. Other specific bottlenecks of the algorithm are also being experimentally analyzed such as the simple-ranking fairness index computation in which the prediction and update of the fairness index is heavily simplified.

The near future next steps also include exploring more in depth the opportunities that power-control offers in terms of efficiency and analyzing the possibilities that the new standards are offering in this sense, and center some work on the synchronization between the uplink and the downlink transmissions. During all the thesis only downlink has been considered but for the final implementation both sides should be clearly covered.

Finally, as also stated in the beginning of the document, the objective for the final solution, is having the wireless hypervisor as only an executor of the network hypervisor decisions, which will be taken from a controller point of view. Performing the optimization from the controller will allow the incorporation of base station collaboration mechanisms and other advanced and complex techniques which can take advantage of the hole wide view of the network.

4.3 Future Research Challenges

As the technology keeps evolving, other new challenges may arise from the near future new developments, challenges that may be interesting to investigate in a near future. The new research efforts trying to propose some new non orthogonal medium access techniques (NOMA) for 5G, with different variants already presented (MUSA, LDS, SCMA and PDMA), which basically improve OFDM assigning every subcarrier to different users at the same time, distinguishing them by allocating different power, different code and different patterns, will present new horizons for the research community when considering future solutions for wireless virtualization.

In the same way as MiMO has started to be investigated, other improvements such as mmWaves, massiveMIMO and beamforming could also be interesting to consider when applying wireless virtualization.

Finally, carrier aggregation, introduced in LTE Advanced and which will gain importance in 5G (allowing signaling information to be transmitted through a different frequency band from the data information transmission) will also give fresh air to the research community, opening a new path to take into account. Options like using mmWaves just to transmit control signaling while keeping the data at a 3G frequency band, achieving this way more useful bandwidth make the use of carrier aggregation very attractive for future proposals.

In any of the cases, the flexibility with which the solution proposed has been designed and the key implementation issues discussed throughput this thesis, will make the adoption of these forthcoming next generation technologies much more straightforward to realize.

REFERENCES

[1] Akyildiz, I. F., Wang, P., and Lin, S.-C., "Softair: A software defined networking architecture for 5g wireless systems," *Computer Networks*, vol. 85, pp. 1–18, 2015.

[2] Andrews, J. G., Buzzi, S., Choi, W., Hanly, S. V., Lozano, A., Soong, A. C., and Zhang, J. C., "What will 5g be?," *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 6, pp. 1065–1082, 2014.

[3] Arora, A., Yoon, S.-G., Choi, Y.-J., and Bahk, S., "Adaptive txop allocation based on channel conditions and traffic requirements in ieee 802.11 e networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 3, pp. 1087–1099, 2010.

[4] Bianchi, G., "Performance analysis of the ieee 802.11 distributed coordination function," *IEEE Journal on selected areas in communications*, vol. 18, no. 3, pp. 535–547, 2000.

[5] Capozzi, F., Piro, G., Grieco, L. A., Boggia, G., and Camarda, P., "Downlink packet scheduling in lte cellular networks: Key design issues and a survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 2, pp. 678–700, 2013.

[6] Cooklev, T., *Wireless Communication Standards: A Study of IEEE 802.11, 802.15, 802.16*. IEEE Standards Association, 2004.

[7] Gast, M. S., *802.11 ac: A survival guide*. " O'Reilly Media, Inc.", 2013.

[8] Liang, C. and Yu, F. R., "Wireless network virtualization: A survey, some research issues and challenges," *Communications Surveys & Tutorials, IEEE*, vol. 17, no. 1, pp. 358–380, 2015.

[9] Lu, X., Yang, K., Liu, Y., Zhou, D., and Liu, S., "An elastic resource allocation algorithm enabling wireless network virtualization," *Wireless Communications and Mobile Computing*, vol. 15, no. 2, pp. 295–308, 2015.

[10] Lu, X., Yang, K., and Zhang, H., "An elastic sub-carrier and power allocation algorithm enabling wireless network virtualization," *Wireless personal communications*, vol. 75, no. 4, pp. 1827–1849, 2014.

[11] Mehlführer, C., Ikuno, J. C., Simko, M., Schwarz, S., Wrulich, M., and Rupp, M., "The vienna lte simulators-enabling reproducibility in wireless communications research.," *EURASIP J. Adv. Sig. Proc.*, vol. 2011, p. 29, 2011.

[12] Oliver, A. C., Stampoultzis, G., Sengupta, A., Klute, R., and Fisher, D., "Apache poi-the java api for microsoft documents," 2002.

[13] Perahia, E. and Stacey, R., *Next Generation Wireless LANS: 802.11 n and 802.11 ac*. Cambridge university press, 2013.

[14] Sesia, S., Toufik, I., and Baker, M., *LTE-the UMTS long term evolution*. Wiley Online Library, 2015.

[15] Wen, H., Tiwary, P. K., and Le-Ngoc, T., "Current trends and perspectives in wireless virtualization," in *Mobile and Wireless Networking (MoWNeT), 2013 International Conference on Selected Topics in*, pp. 62–67, IEEE, 2013.

[16] Wen, H., Tiwary, P. K., and Le-Ngoc, T., "Wireless virtualization," in *Wireless Virtualization*, pp. 41–81, Springer, 2013.

[17] Xiao, Y., "Ieee 802.11 e: Qos provisioning at the mac layer," *IEEE Wireless Communications*, vol. 11, no. 3, pp. 72–79, 2004.

[18] Zaki, Y., Zhao, L., Goerg, C., and Timm-Giel, A., "Lte wireless virtualization and spectrum management," in *Wireless and Mobile Networking Conference (WMNC), 2010 Third Joint IFIP*, pp. 1–6, IEEE, 2010.

[19] Zaki, Y., Zhao, L., Goerg, C., and Timm-Giel, A., "A novel lte wireless virtualization framework," in *Mobile Networks and Management*, pp. 245–257, Springer, 2010.

Towards Wireless Virtualization for
5G Cellular Systems

Albert Gran Alcoz

74 Pages

Directed by Professor Ian F. Akyildiz

Although it has been defined as one of the most promising key enabling technologies for the forthcoming fifth generation cellular networks, wireless virtualization still has several challenges remaining to be addressed. Amongst those, resource allocation, which decides how to embed the different wireless virtual networks on the physical relying infrastructure, is the one receiving maximum attention. This project aims at finding the optimal resource allocation for each virtual network, in terms of channel resources, power levels and radio access technologies so that the data rate requested by each virtual network can be guaranteed and the global throughput efficiency can be maximized.